PERIODICAL ROOM
GENERAL LIBRARY
UNIV. OF MICH.

DEC 22 1932

VOL. XXIII DECEMBER 1932 NO. 9

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

CONTENTS

Title-page and Index for Volume XXIII (1932) . . . . . i–viii

Twin Similarities in Occupational Interests. 641
HAROLD D. CARTER

The Reliability and Validity of Multiple-response Tests When Presented Orally. 656
V. M. SIMS AND L. B. KNOX

Auditorium Versus Classroom Showing of Motion Pictures in History Teaching. 663
DANIEL C. KNOWLTON AND J. WARREN TILTON

The Validity of Personality Ratings of Preschool Children. 671
HERBERT S. CONRAD

Scoring Objective Tests 681
NOEL B. CUFF

A Study of High School Graduates with Reference to Level of Intelligence. 687
CONSTANCE E. DOWD

A Diagnostic Study of Indian Handwriting. 703
ELIZABETH WEISSER

Book Reviews 708

$5.00 per Year · Published Monthly Except June to August by

WARWICK & YORK, INC.

BALTIMORE, MD.

Entered as Second Class Matter Nov. 15, 1921, at the Post Office at Baltimore, Md. under the Act of March 3, 1879; additional entry as second class matter at York, Pa.

THE JOURNAL OF EDUCATIONAL PSYCHOLOGY

ESTABLISHED 1910

Board of Editors

HAROLD RUGG

J. CARLETON BELL ARTHUR I. GATES RUDOLF PINTNER E. L. THORNDIKE
FRANK N. FREEMAN V. A. C. HENMON L. M. TERMAN GOODWIN B. WATSON

The price of the Journal is $5.00 a year in the United States; $5.30 in foreign countries. Subscriptions for less than a year are charged at the rate of 70 cents for each number ordered. Back volumes are $6.00 each; back issues, 90 cents each except when three or more years old, and then $1.00 each.

Subscribers should notify the publishers of change in address at least four weeks before publication of the issue with which the change is to take effect. Claims for non-receipt of an issue will not be honored unless made within two weeks after receipt of the next succeeding number.

Manuscripts for publication, books or other materials for review, and correspondence should be addressed to the publishers.

WARWICK AND YORK, INC., *Baltimore, Maryland*

Please Let Us Know—
Do You Receive
"WORD STUDY"?

Throughout the school year we issue periodically, for the benefit of English teachers, a publication entitled WORD STUDY in which are presented many interesting and helpful ideas relating to the teaching of the subject. It will help us to know whether you receive WORD STUDY; if you do, what your opinion of it is; and if you do not, whether you would like to receive it free of expense. We will appreciate a word from you. In writing please mention

The Merriam-Webster *is the* "Supreme Authority"

One hundred years of experience insure its accuracy and scholarship and have established it as the authority in American schools.

WEBSTER'S NEW INTERNATIONAL DICTIONARY

The Presidents and Department Heads of leading Universities and Colleges give their indorsement. Write for Free illustrated pamphlet with sample pages.

G. & C. MERRIAM CO.
Springfield, Mass.

TEACHERS' HEADQUARTERS

European
Plan

Entirely
Fire-Proof

Centrally
Located

EDWARD DAVIS
Manager

HOTEL RENNERT
BALTIMORE, MARYLAND

The Psychology of Learning
1928 Revision By WILLIAM H. PYLE
$2.20+12¢ postage
Warwick & York, Inc., *Baltimore, Maryland*

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume XXIII | December, 1932 | Number 9

TWIN SIMILARITIES IN OCCUPATIONAL INTERESTS*

HAROLD D. CARTER

Stanford University

Some very important problems in the analysis of occupational interests may be approached through study of family resemblances in those traits which the interest scales measure. In his recent book, Fryer[2] furnishes an admirable summary of literature bearing on this field of mental measurement, and his discussion of the inheritance of interests shows clearly the need for data of the type reported in the present paper. He points out that psychologists commonly regard interests as developed through experience and environmental in cause. He gives the further observation that this conclusion is a negative one, based on the absence of proof that interests are of inherited origin, and states that the positive proof that interests are determined to a large degree by the chance stimulations of the environment is not forthcoming, and that no work has been done upon the problem. It seems that knowledge concerning the origins of interests should be of value to the psychologist intent upon improving and perfecting the measuring instruments, to the vocational counselor studying the problems of the individual, and to the student of mental heredity in his analysis of the relations among interests, abilities, and experiences. In view of these considerations it has been considered worth while to report the present findings concerning the similarities of twins in occupational interests. In determining the origins of the traits measured further studies are of course necessary, but it is felt that

* The present study represents one aspect of the intensive investigation of twin similarities which the writer is carrying out at Stanford University under the direction of Dr. Lewis M. Terman. These studies were made possible by a grant from the Social Science Research Council.

determination of the degree of resemblance of related persons is an important first step in the undertaking.

THE DATA

The Strong Vocational Interest Blank was administered to one hundred twenty sets of twins, and the blanks were scored for the interests characteristic of successful persons in twenty-three different occupations. Table I shows some facts concerning the subjects, the groups into which they fall, and the age groupings.

TABLE I.—THE SUBJECTS

	Numbers of pairs		
	Junior and senior high school group	Older group not in school	Total No. of pairs
Monozygotic twins			
Males	17	4	21
Females	12	10	22
Like-sex dizygotic twins			
Males	22	0	22
Females	21	0	21
Dizygotic twins of unlike sex	33	1	34
Total No. of pairs	105	15	120

From Table I it may be seen that most of the subjects are in the junior and senior high schools. It is here assumed that the inclusion of a few older pairs has not materially affected the particular results reported here. Further treatment will be reported to show exactly what effect the age factor may have had upon the correlations. It is generally assumed that interests change with age most rapidly during youth, hence an inquiry has been made concerning age effects using the younger group. The present data are not adequate for determining the differences between the younger and older groups in similarity of interests, as with such small numbers of cases the errors of sampling are large. Inspection of the data suggests that older twins do not differ significantly from younger twins in similarity of interests, but perhaps in general the older pairs are slightly less alike. However, scores of older persons may tend to be slightly higher on some of the scales and lower on others, and hence because of the effect upon

the range and central tendencies of distributions of scores, the inclusion of the few older pairs may have had some effect upon the correlations reported. That this effect is probably a slight one will be pointed out later.

Classification of the twins into identical and fraternal groups was made according to the criteria outlined by Dahlberg,[1] Siemens,[5] Newman,[4] Muller,[3] and others.

THE EFFECTS OF AGE

Table II shows the effects of age upon interest test scores for the younger group in junior and senior high school. There were one hundred ten boys and one hundred ten girls in the groups for which the correlations were calculated, hence the probable error of each correlation in Table II is approximately .06. Preliminary inspection of the data had suggested that the group was sufficiently homogeneous so that it would not be necessary to subdivide it according to age for further study. The data of Table II support this conclusion, since there appears to be no significant correlation of interest scores with age for this group. Since it is generally assumed that the change of interests with age is greatest for the younger ages, it seems unlikely that separate treatment need be given to the few older pairs who are not included in the school group.

Preliminary study has shown that there are sex differences in means and standard deviations of the distributions of scores on a number of the occupational interest scales. These sex differences are sufficiently large to be demonstrably significant, and hence must be taken into account. This paper deals primarily with twin similarities, and hence does not attempt an exhaustive analysis of the factors underlying those sex differences. Such analysis is reserved for another paper, and the correlations reported here are calculated separately for the two sexes to avoid spurious correlation which might arise from combination of groups differing in standard deviation and means.

The Strong Vocational Interest Test which has been used in the present study has of course been developed mainly for use with older subjects, but from several points of view the present data are especially valuable. The importance of genetic studies to find out all we can about the origins and development of occupational interests is not questioned. The findings here reported do not depend in any way upon the norms established for older groups to determine diagnostic ratings. The administering of the tests was carefully supervised by

TABLE II.—SHOWING THE CORRELATIONS OF AGE WITH INTEREST TEST SCORES, FOR BOYS AND GIRLS SEPARATELY. ALL ARE JUNIOR AND SENIOR HIGH SCHOOL STUDENTS. THERE WERE ONE HUNDRED TEN BOYS RANGING IN AGE FROM TWELVE TO NINETEEN YEARS INCLUSIVE, WITH A STANDARD DEVIATION OF 1.5 YEARS, AND THE MEAN AT AGE 15.9. THE ONE HUNDRED TEN GIRLS RANGED IN AGE FROM THIRTEEN TO EIGHTEEN YEARS INCLUSIVE, WITH A STANDARD DEVIATION OF 1.25 YEARS, AND A MEAN AT AGE 16.1

	Correlations	
	For boys	For girls
1. Advertiser	−.04	.01
2. Architect	.08	−.10
3. Artist	−.08	−.06
4. Certified public accountant	.08	−.17
5. Chemist	.10	−.10
6. Doctor	−.02	−.16
7. Engineer	.13	−.15
8. Farmer	.22	−.20
9. Journalist	−.06	−.06
10. Lawyer	.07	−.09
11. Life insurance salesman	.00	.10
12. Minister	−.03	.07
13. Personnel manager	.11	.16
14. Psychologist	−.02	−.12
15. Purchasing agent	.25	.16
16. Real estate salesman	.08	.07
17. School man	.09	−.02
18. Vacuum cleaner salesman	−.01	.14
19. Y.M.C.A. secretary	.02	.12
20. Office clerk	.12	.12
21. Physicist	.00	−.26
22. Mathematician	.02	−.30
23. City school superintendent	.18	.03
Mean of correlations	.06	−.035
Mean of absolute magnitude of these correlations	.08	.12

Note: It is unnecessary to report each probable error for correlations of this narrow range of magnitude. The largest correlation is .30, and its probable error is .058. The smallest correlation is .00 and its probable error is .064. Hence the probable error of each coefficient, given to two places, is .06.

the writer or by an equally competent examiner, to eliminate possible misunderstandings on the part of the subjects. Because of the importance of the twin control in the determination of the inheritance of interests, and because the group here investigated comprise a more nearly random sampling of the population than could be obtained using older subjects, the data are doubly valuable for the purposes for which they were obtained.

The literature on interest inventories shows that the interest test measures something which does not depend primarily upon intelligence, but that scoring schemes can be devised which will yield ratings of intelligence and scholarship. For this reason it has been thought desirable to report the correlations between intelligence and scores on each of the interest scales for the present groups. Those correlations are given in Table III.

The work of Strong,[6] and that of Thurstone[8] have shown separately by different types of analysis that the interests measured by these different scales tend to fall into related groups. There are four such groups, depending upon interest in science, interest in language, interest in people, and interest in business. Table IV presents the correlations of intelligence and interests for the scales grouped according to these four types of interest.

The data of Table III and of Table IV indicate that there is a definite correlation between intelligence and certain types of interest, but that the relationship is not a close one. As might be expected, the relationships are not the same for boys as for girls. The interests in science are positively correlated with intelligence for both groups, but the correlation is greater for females. In general, interests in language are uncorrelated with intelligence for boys, but there is a definite positive correlation for girls. It appears that for this group of girls, interests in people show a consistent slight negative correlation with intelligence, while for boys there is a small but consistent positive correlation. The business group of interests correlate negatively with intelligence for both groups, but the relationship is closer in the case of the females. The important facts are that intelligence is definitely related to certain types of interest, that the relationships show definite sex differences, and that the correlations are low enough to show that the interest scales measure traits not primarily associated with intelligence. It is necessary to point out that the subjects were all junior and senior high school students, and that different results might be obtained for different groups.

TABLE III.—SHOWING THE CORRELATIONS OF TERMAN GROUP TEST IQ'S WITH OCCUPATIONAL INTEREST SCORES, FOR BOYS AND GIRLS SEPARATELY. THERE WERE SEVENTY BOYS, WHOSE IQ'S RANGED FROM SEVENTY-THREE TO ONE HUNDRED FORTY, WITH A MEAN OF 101.1 AND A STANDARD DEVIATION OF 14.4 POINTS. THERE WERE SEVENTY-FIVE GIRLS, WHOSE IQ'S RANGED FROM SEVENTY-THREE TO ONE HUNDRED TWENTY-SEVEN, WITH A MEAN OF 101.5 AND A STANDARD DEVIATION OF 12.8 POINTS. ALL THE SUBJECTS WERE JUNIOR AND SENIOR HIGH SCHOOL STUDENTS

	Correlation coefficients			
	For the boys		For the girls	
	r	PE	*r*	PE
1. Advertiser	−.11	.08	.23	.07
2. Architect	.07	.08	.30	.07
3. Artist	.04	.08	.24	.07
4. C.P.A.	.09	.08	−.08	.08
5. Chemist	.31	.07	.36	.07
6. Doctor	.04	.08	.32	.07
7. Engineer	.17	.08	.36	.07
8. Farmer	.02	.08	.20	.07
9. Journalist	−.09	.08	.22	.07
10. Lawyer	.13	.08	.31	.07
11. Life insurance salesman	−.19	.08	.01	.08
12. Minister	.23	.08	−.08	.08
13. Personnel manager	.27	.07	−.03	.08
14. Psychologist	.36	.07	.34	.07
15. Purchasing agent	.00	.08	−.13	.08
16. Real estate salesman	−.15	.08	.05	.08
17. School man	.35	.07	−.04	.08
18. Vacuum cleaner salesman	−.14	.08	−.44	.06
19. Y.M.C.A. secretary	.14	.08	−.19	.07
20. Office clerk	.09	.08	−.24	.07
21. Physicist	.19	.08	.33	.07
22. Mathematician	.18	.08	.23	.07
23. City school superintendent	.32	.07	.03	.08
Mean of correlations	.10	...	.10	
Mean of absolute values of these correlations	.16	...	.21	

TABLE IV.—THE RELATION OF TYPE OF INTEREST TO INTELLIGENCE
(The correlations are taken from Table III)

	Correlations	
	For boys	For girls
The "science" group of interests:		
1. Psychologist	.36	.34
2. Mathematician	.18	.23
3. Physician	.04	.32
4. Chemist	.31	.36
5. Physicist	.19	.33
6. Architect	.07	.30
7. Engineer	.17	.36
8. Farmer	.02	.20
Mean of correlations	.17	.30
The "language" group of interests:		
1. Journalist	−.09	.22
2. Lawyer	.13	.31
3. Advertiser	−.11	.23
4. Artist	.04	.24
Mean of correlations	−.01	.25
The "people" group of interests:		
1. Y.M.C.A. secretary	.14	−.19
2. City school superintendent	.32	.03
3. Personnel manager	.27	−.03
4. Minister	.23	−.08
5. Teacher	.35	−.04
Mean of correlations	.26	−.06
The "business" group of interests:		
1. Vacuum cleaner salesman	−.14	−.44
2. Real estate salesman	−.15	.05
3. Life insurance salesman	−.19	.01
4. Purchasing agent	.00	−.13
5. Office clerk	.09	−.24
6. C.P.A	.09	−.08
Mean of correlations	−.05	−.14

TWIN RESEMBLANCES

Tables V and VI show the correlations which were obtained for all groups of twins, for each of the twenty-three interest scales. For

TABLE V.— SHOWING THE CORRELATIONS OF SCORES ON THE STRONG VOCATIONAL INTEREST BLANK FOR MALE LIKE-SEX TWINS, AND FOR UNLIKE-SEX TWINS

	Identical twins, twenty-one pairs, all males		Fraternal like-sex twins, twenty-two pairs, all males		Unlike-sex twins, thirty-four pairs	
	r	PE	*r*	PE	*r*	PE
1. Advertiser	.39	.13	.23	.14	.12	.11
2. Architect	.71	.08	.06	.15	.02	.11
3. Artist	.35	.13	.02	.15	.08	.11
4. C.P.A.	.48	.12	.24	.14	.22	.11
5. Chemist	.74	.07	.29	.14	.46	.09
6. Doctor	.59	.10	.00	.15	.46	.09
7. Engineer	.84	.04	.22	.14	.44	.09
8. Farmer	.56	.10	.64	.09	.14	.11
9. Journalist	.47	.12	.24	.14	.04	.11
10. Lawyer	.77	.06	.18	.14	.27	.11
11. Life insurance salesman	.60	.10	.24	.14	.51	.08
12. Minister	.58	.10	.13	.15	.32	.10
13. Personnel manager	.37	.13	.27	.14	.21	.11
14. Psychologist	.28	.14	.23	.14	.47	.09
15. Purchasing agent	.52	.11	.28	.14	.36	.10
16. Real estate salesman	.33	.13	.05	.15	.28	.10
17. School man	.27	.14	.33	.13	.27	.11
18. Vacuum cleaner salesman	.66	.08	.14	.15	.37	.11
19. Y.M.C.A. secretary	.64	.09	.16	.15	.37	.10
20. Office clerk	.60	.10	.10	.15	.41	.10
21. Physicist	.79	.06	.07	.15	.48	.09
22. Mathematician	.56	.10	.14	.15	.29	.10
23. City school superintendent	.59	.10	.42	.12	.33	.10
Mean of correlations	.55	...	.20	...	.30	

the like-sex pairs of twins a double-entry method was used in making the scatterplots from which the correlations were calculated. For unlike-sex pairs a single-entry system was used, the boys scores being put on one axis and the girls on the other axis. The correlations for

TABLE VI.—SHOWING THE CORRELATIONS OF SCORES ON THE STRONG VOCATIONAL INTEREST BLANK FOR FEMALE LIKE-SEX TWINS, AND FOR UNLIKE-SEX TWINS

	Identical twins, twenty-two pairs, all females		Fraternal like-sex twins, twenty-one pairs, all females		Unlike-sex twins, thirty-four pairs
	r	PE	*r*	PE	*r*
1. Advertiser	.23	.14	.35	.13	.12
2. Architect	.52	.11	.45	.12	.02
3. Artist	.52	.11	.58	.10	.08
4. C.P.A	.46	.12	.38	.13	.22
5. Chemist	.54	.11	.37	.13	.46
6. Doctor	.59	.10	.58	.10	.46
7. Engineer	.63	.09	.06	.15	.44
8. Farmer	.42	.12	.26	.14	.14
9. Journalist	.22	.14	.27	.14	.04
10. Lawyer	.23	.14	.17	.15	.27
11. Life insurance salesman	.47	.12	.38	.13	.51
12. Minister	.58	.10	.13	.15	.32
13. Personnel manager	.54	.11	.42	.12	.21
14. Psychologist	.44	.12	.18	.14	.47
15. Purchasing agent	.20	.14	.42	.12	.36
16. Real estate salesman	.28	.14	.14	.15	.28
17. School man	.58	.10	.16	.15	.27
18. Vacuum cleaner salesman	.42	.12	.58	.10	.37
19. Y.M.C.A. secretary	.51	.11	.36	.13	.37
20. Office clerk	.32	.13	.31	.14	.41
21. Physicist	.51	.11	.37	.13	.48
22. Mathematician	.52	.11	.37	.13	.29
23. City school superintendent	.66	.08	.16	.15	.33
Mean of correlations	.45	...	.32	...	.30

unlike-sex twins are entered in both Table V and Table VI, for purposes of comparison.

Tables VII and VIII show further comparisons of the different groups of twins, giving average correlations based on the data of Tables V and VI. These comparisons were considered desirable

TABLE VII.—SHOWING COMPARISON OF CORRELATIONS OF SCORES ON THE STRONG VOCATIONAL INTEREST BLANK, FOR THREE GROUPS OF TWINS. THE CORRELATIONS WERE CALCULATED SEPARATELY FOR MALES AND FOR FEMALES OF THE LIKE-SEX GROUPS, AS SHOWN IN TABLES V AND VI. THE AVERAGES ARE PRESENTED HERE

	Identical twins, forty-three pairs	Like-sex fraternal twins, forty-three pairs	Unlike-sex twins, thirty-four pairs
1. Advertiser	.31	.29	.12
2. Architect	.62	.26	.02
3. Artist	.44	.30	.08
4. C.P.A.	.47	.31	.22
5. Chemist	.64	.33	.46
6. Doctor	.59	.29	.46
7. Engineer	.74	.14	.44
8. Farmer	.49	.45	.14
9. Journalist	.34	.26	.04
10. Lawyer	.50	.18	.27
11. Life insurance salesman	.54	.31	.51
12. Minister	.58	.13	.32
13. Personnel manager	.46	.34	.21
14. Psychologist	.36	.20	.47
15. Purchasing agent	.36	.35	.36
16. Real estate salesman	.30	.10	.28
17. School man	.42	.24	.27
18. Vacuum cleaner salesman	.54	.36	.37
19. Y.M.C.A. secretary	.58	.26	.37
20. Office clerk	.46	.20	.41
21. Physicist	.65	.22	.48
22. Mathematician	.54	.26	.29
23. City school superintendent	.62	.29	.33
Mean of correlations	.50	.26	.30

TABLE VIII.—COMPARISON OF THE CORRELATIONS OF SCORES ON THE STRONG VOCATIONAL INTEREST SCALES, FOR IDENTICAL TWINS AND FOR FRATERNAL TWINS. SINCE THE PREVIOUS ANALYSIS SHOWED THE LIKE-SEX FRATERNAL PAIRS NOT MORE SIMILAR THAN THE UNLIKE-SEX PAIRS, THE TWO GROUPS OF FRATERNAL TWINS ARE COMBINED HERE. THE COEFFICIENTS IN THIS TABLE ARE WEIGHTED AVERAGES, SINCE THE NUMBERS OF CASES IN THE TWO GROUPS DIFFERED

	Identical twins, forty-three pairs	Fraternal twins, average correlation for like and unlike-sex pairs. Total of seventy-seven pairs
1. Advertiser	.31	.21
2. Architect	.62	.15
3. Artist	.44	.20
4. C.P.A.	.47	.27
5. Chemist	.64	.39
6. Doctor	.59	.36
7. Engineer	.74	.27
8. Farmer	.49	.31
9. Journalist	.34	.16
10. Lawyer	.50	.22
11. Life insurance salesman	.54	.40
12. Minister	.58	.21
13. Personnel manager	.46	.28
14. Psychologist	.36	.32
15. Purchasing agent	.36	.35
16. Real estate salesman	.30	.18
17. School man	.42	.25
18. Vacuum cleaner salesman	.54	.36
19. Y.M.C.A. secretary	.58	.31
20. Office clerk	.46	.29
21. Physicist	.65	.33
22. Mathematician	.54	.27
23. City school superintendent	.62	.31
Mean of correlations	.50	.28

because it was felt that much of the variation in magnitude of individual correlation coefficients is due to the smallness of the populations.

These further comparisons give an idea as to what might be the result if larger groups of subjects were tested.

It is evident from consideration of the trends of results shown in Tables V, VI, VII, and VIII that the interests of identical twins tend to be more similar than those of like sex fraternal twins, for both male and female pairs. There is no evidence that like-sex fraternal pairs are any more similar than unlike-sex pairs. The differences in similarity between identical and fraternal pairs are greater for males than for females.

NATURE AND NURTURE INFLUENCES

It is to be expected that members of the same family will show some degree of resemblance in occupational interests, regardless of one's point of view on the nature-nurture question. Unlike-sex pairs of twins will show some resemblance because they are siblings; they are somewhat similar in heredity, and their environments are certainly more alike than those of randomly selected pairs. Like-sex fraternal twins will also be similar, for the same two reasons, and if environment be important they may be more similar than are unlike-sex pairs. Moreover, if those aspects of inherited constitution which are determined by genes in the sex chromosome are important, all like-sex pairs will be more similar than unlike-sex pairs. The identical twins are expected to show greater similarity than the fraternal pairs, because of more similar heredity, and perhaps also because of more similar environments. It is likely that both heredity and environment have an effect upon family resemblances in interests of the type here considered. Determination of the extent of those resemblances is a necessary first step, and further observations are necessary to find out the relative contributions of nature and of nurture. With these *a priori* considerations in mind, we may now turn to the specific results of the present investigation.

CONCLUSIONS AND SUGGESTIONS

Since the analysis here reported is based upon study of a small number of subjects, it is necessary to be cautious in drawing conclusions. The fact that correlations tend to be higher for all groups on some of the scales than for other scales may be partly the result of differences in reliability of the various scales, but it does not seem appropriate to correct the coefficients for attenuation until reliability

coefficients can be obtained which are based on comparable groups. Those reported in the literature are based on older subjects.

In spite of various objections which may be raised, study of the available data seems to justify the following tentative conclusions:

1. There are definite family resemblances in occupational interests, as measured by the various scales of the Strong Vocational Interest Blank. The extent of these resemblances as indicated by the present data is represented by average coefficients of correlation of .50 for identical twins, and of .28 for fraternal twins.

2. Since it is certain that the interests of the monozygotic twins are more similar than those of the fraternal twins, and since the data suggest that there is no greater similarity of like-sex fraternal twins than of unlike-sex twins, it seems probable that hereditary factors are more important in determining interests than are environmental factors. This conclusion can only be made tentatively, however, since a part of the greater similarity of identical twins must be attributed to greater similarity of environment. The present writer believes that such greater similarity of environment is a fact which must be taken into consideration, but that it is a far less important fact than the greater similarity of heredity.

3. The present finding that male like-sex fraternal twins are, for this sampling at least, less similar than unlike-sex twins, raises several important questions. The finding is surprising in view of the common assumption that the interests of girls tend to be much different from those of boys. When all the data on male and female pairs are considered, a possible explanation is suggested. Girls may show less variability and less individuality than do boys, where occupational interests are concerned, and male fraternal twins may be in general less closely associated than are female fraternal twins, or even unlike-sex twins.

4. The greater similarity of all pairs on some of the interest scales may be partly the result of greater reliability of some of the scales, and also may be partly the result of differences in the specificity of the traits measured by those scales. It is also probable that the correlations are lower for those traits more affected by environmental influences. Further analysis is desirable, in order to answer these questions.

5. It is noteworthy that in these traits measured by these twenty-three scales, traits which there is every reason to believe are more affected by environmental factors than are such traits as general intelli-

gence, the correlations for related pairs of individuals are definitely lower than for most intellectual traits. This may be in part due to the unreliability of the measuring instruments, for although fair reliability has been reported for most of the scales, it must be kept in mind that we are here dealing with a younger group than that for which the test was devised. If the reliabilities are lower for such groups as this, it is probably an indication that environmental influences *do* affect the development of these interests. Developmental studies are necessary to supplement studies of the type here presented.

6. In view of previous literature on the subject, it is interesting to note that there seems to be no close relation between scores on these scales and chronological age, during the junior and senior high school period. That is, the correlations of scores with age for the age range from twelve to twenty are by no means so great as they would be for a trait such as general intelligence, or vocabulary, or ability in arithmetic. On the contrary, they approximate zero in magnitude. This finding may be partly explained by the construction of the test, the method of administering it, and the method of scoring. It is certain that there are great individual differences in scores at these young ages, and that they are not due to chronological age, nor in any way closely related to chronological age. It is admitted of course that the scores may be less valid as measures of vocational interests at this age level, but other considerations suggest that they measure stable traits of the individual.

7. The scores on the test are definitely related to intelligence test scores, but the correlations are not high enough to account for the findings regarding twin similarities. For some of the scales the correlations are negative, for some positive, and for some zero. These correlations are systematically related to type of interest, and they differ for boys and girls. Because of the lesser vocational information of these younger subjects, it is likely that these findings might differ somewhat from those which would be obtained for older groups.

8. It is important for the student of heredity to know that the greater similarity of monozygotic than dizygotic twins extends to traits such as these, and it is also important that the vocational counselor should know the facts concerning family resemblances in occupational interests. The present study may be considered useful as a preliminary investigation of the development of occupational interests.

REFERENCES

1. Dahlberg, Gunnar: "Twin Births and Twins from a Hereditary Point of View." Bokforlags A-B. Tidens Tryckeri, Stockholm, 1926, pp. 296 + 100.
2. Fryer, Douglas: "The Measurement of Interests." Henry Holt and Company, New York, 1931, pp. XXXVI + 488.
3. Muller, H. J.: Mental Traits and Heredity, *Journal of Heredity*, Vol. 16, 1925, pp. 433–448.
4. Newman, H. H.: Studies of Human Twins. Part I, Methods of Diagnosing Monozygotic and Dizygotic Twins, *Biological Bulletin*, Vol. 55, 1928, pp. 287–297. Part II, Asymmetry Reversal or Mirror Imaging in Identical Twins, *Ibid.*, 55, pp. 298–315.
5. Siemens, H. W.: The Diagnosis of Identity in Twins, *Journal of Heredity*, Vol. 18, 1927, pp. 201–209.
6. Strong, E. K., Jr.: "Manual for the Vocational Interest Blank." Stanford University Press.
7. Strong, E. K., Jr.: "Change of Interests with Age." Stanford University Press, 1931, pp. XIX + 235.
8. Thurstone, L. L.: A Multiple Factor Study of Vocational Interests, *Personnel Journal*, Vol. X., Oct., 1931, pp. 198–205.

THE RELIABILITY AND VALIDITY OF MULTIPLE-RESPONSE TESTS WHEN PRESENTED ORALLY

V. M. SIMS AND L. B. KNOX

University of Alabama

One of the most common objections raised to the use of objective tests for measuring the results of instruction is the large amount of time, equipment, and stenographic assistance necessary for duplicating tests. Not only are adequate duplicating facilities often lacking; but many teachers working in systems where such assistance is supposedly rendered complain that at examination times there is usually such a rush that only the more fortunate—or farsighted—get their examinations prepared in time for use. From an administrative standpoint the cost of paper, stencils, and clerical assistance for a large school system must be a respectable item in the annual budget.

The only apparent justification for such wholesale duplication of tests is that they cannot be successfully administered in any other manner. There seems to be a more or less common assumption that objective tests must be presented visually if they are to be satisfactory instruments of measurement. The experimental evidence available to date does not support this assumption. Chapman, as long ago as 1920, presented an intelligence test which was designed to be administered orally, the child writing the responses on blank sheets of paper.[1] This examination consisted of one hundred items of the recall type. A revised edition, published in 1925, was found to have a reliability coefficient of .84, determined by repeating the test within the week to a group of eighth grade children; and the correlation between Chapman's test and the *Terman Group Test of Mental Ability* was .87 ($N = 62$).[2] His results would seem to indicate that simple recall items give a fairly valid and reliable measure when presented orally.

Ross and Gard,[3] after an extensive investigation of the validity and reliability of dictating the *Otis Self-administering Test of Mental*

---

[1] Chapman, J. C.: A Group Intelligence Examination without Prepared Blanks. *Journal of Educational Research*, December, 1920, pp. 777–786.

[2] Chapman, J. C.: A Group Intelligence Examination without Prepared Blanks, Revised Form. *Journal of Educational Research*, April, 1925, pp. 269–279.

[3] Ross, C. C. and P. D. Gard: Two Modified Methods of Administering Two Standardized Group Intelligence Tests. *Bulletin of the Bureau of School Service*, College of Education, University of Kentucky, Vol. II, No. 4.

Ability and *Terman's Group Test of Mental Ability*, concluded that the oral method was almost as satisfactory as the visual method. Both of these tests are made up of several different types of objective items.

Lehman[1] and Jenson[2] have recently reported comparisons of true-false examinations presented orally with the same examination presented in mimeographed form. Lehman found an average correlation of .51 between odds and evens in an eighty-five item true-false examination which was read to the classes, each item being read twice; and the same test presented immediately afterward in mimeographed form gave an average correlation between odds and evens of .47. He also reported correlations between the two that were as high as one would expect from tests with such low reliability. Oral-odds with mimeographed-evens gave an average correlation of .44 and oral-evens with mimeographed-odds gave an average of .49. Jenson found an average reliability for a visual true-false test of .77 and the same test presented orally gave a coefficient of .79. These experiments would lead one to conclude that the true-false test is just as good a measure when presented orally as when presented visually.

The study here reported is concerned with the reliability and validity of orally presented multiple-response tests. Comparisons of the reliability of a three-response, four-response and five-response test presented orally are made and each type is further compared with another form of the same test presented in printed form.

Four forms of the *Thorndike Test of Word Knowledge*, a one-hundred item five-response test, were administered to a group of one hundred students enrolled in grades nine through twelve. The four forms were all administered within a period of one month. Form A (known hereafter as the visual 5-*R* test), given first, was administered in the usual manner, Thorndike's directions being observed. For Form B, the second test administered, (known as the oral 3-*R* test) two of the wrong choices given for each test word were eliminated and the test given orally as a three-response test. One of the wrong choices given for each test word in Form C was eliminated, and this test, the oral 4-*R* test, was given orally for the third test. Form D,

[1] Lehman, H. C.: The Oral Versus the Mimeographed True-false. *School and Society*, October 5, 1929, pp. 470–472.

[2] Jenson, M. B.: An Evaluation of Three Methods of Presenting True-false Examinations; Visual, Oral and Visual-oral. *School and Society*, November 19, 1930, pp. 675–677.

administered last, was presented orally with all of the choices as the oral 5-*R* test.

The directions used with the oral tests were the same for the three forms. After giving each student a blank sheet with the numbers from one to one hundred placed in columns the following directions were read: "Listen carefully to the first word that I read, then listen to the other three (four, five) words read. One of the last three (four, five) words means the same or about the same as the first word. Decide which one has the same meaning and write it after the number *one* on your sheet. If you cannot decide after the first reading wait—it will be read again. If you do not get it the second time draw a line after the number *one*. There will be one hundred such words for you to decide upon. Be sure to draw a line for words that you do not know the meaning of." The words were read with a pause between the test word and the response words, and with a longer pause between the two readings. The tests were all administered by the same tester, a person with whom the subjects were familiar.

The working time for each of the forms was as follows:

Type of test	Time
Visual 5-response	38 min. (ave.)
Oral 3-response	43 min.
Oral 4-response	44 min.
Oral 5-response	45 min.

The tests were scored by two methods, first by correcting for guessing by means of the usual formula, and secondly without any correction for guessing, the score being the number of correct responses. The mean score, converted to a comparable basis as suggested by Thorndike in his *Manual of Directions*, and the standard deviation for the one hundred cases, by each of these scoring methods, for each of the four tests is presented in Table I.

As would be expected the scores corrected for guessing are lower than those uncorrected for guessing. When corrected the visual 5-*R* appears to be the easiest form and the oral 4-*R* the most difficult, but for uncorrected scores the oral 3-*R* is the easiest and the oral 5-*R* the most difficult. The differences are not pronounced however, and should not be considered as very significant. For example, the difference between the uncorrected mean of the visual 5-*R* and the oral

TABLE I.—SHOWING MEAN SCORE AND STANDARD DEVIATION FOR FOUR FORMS

	N	Corrected for guessing		Uncorrected for guessing	
		Mean	SD	Mean	SD
Visual 5-*R*	100	57.9	17.0	65.4	13.8
Oral 3-*R*	100	50.8	17.1	67.1	10.6
Oral 4-*R*	100	48.2	15.2	62.9	12.0
Oral 5-*R*	100	51.7	15.3	61.6	13.1

5-*R* (3.8) is not quite five times the probable error of the difference (.83). It seems safe to conclude that oral tests are but slightly, if any, more difficult than visual tests, and that the oral three-response is somewhat easier than the four- or five-response.

The reliability for each form was determined by splitting halves (odds *vs.* evens) and stepping-up the coefficients found by means of the Spearman-Brown formula. Since the time for testing varied slightly for the different methods the coefficients were further reduced to the coefficient for one hour's testing by means of the same formula. These coefficients of reliability for the four forms, when corrected for guessing and when uncorrected for guessing, are presented in Table II.

TABLE II.—SHOWING RELIABILITY COEFFICIENTS FOR FOUR FORMS

	Corrected for guessing			Uncorrected for guessing		
	$r_{½·½}$	r_{total}	$r_{1\ hr.}$	$r_{½·½}$	r_{total}	$r_{1\ hr.}$
Visual 5-*R*	.78 ± .02	.88	.92	.82 ± .02	.90	.94
Oral 3-*R*	.64 ± .02	.78	.83	.70 ± .03	.82	.87
Oral 4-*R*	.88 ± .01	.94	.95	.83 ± .02	.91	.93
Oral 5-*R*	.74 ± .03	.85	.88	.76 ± .02	.86	.89

The facts presented in this table seem to indicate that the uncorrected score is more reliable than the corrected score (in spite of the increased range for the corrected form), and that the oral 4-*R* is the most reliable, the visual 5-*R* next, and the oral 5-*R* and oral 3-*R* next in order. Although the differences are not so pronounced, this order is the same when the tests are reduced to equivalent times of testing. However, the probable errors of the coefficients are of such a size that

one feels justified in concluding that so far as our findings go there are no pronounced differences in the reliability of the different tests.

Assuming the visual five-response technique to be a valid method of measuring vocabulary, and Ruch's work leads one to believe that not only is the multiple-response technique valid, but that a five response is more satisfactory than any of several responses investigated,[1] then the correlation between the visual test and each of the three oral tests might be considered as a measure of the validity of the oral procedure. Table III presents the coefficients found between the visual five-response and each of the oral tests for both corrected and un-corrected scores. Each coefficient has also been corrected for attenuation due to the unreliability of the measures by means of Spearman's formula.

TABLE III.—SHOWING CORRELATION BETWEEN VISUAL FORM AND ORAL FORMS

	Corrected for guessing		Uncorrected for guessing	
	r_{raw}	$r_{cor.}$	r_{raw}	$r_{cor.}$
$r_{visual\text{-}oral\,3\text{-}R}$	.58 ± .04	.71	.54 ± .04	.64
$r_{visual\text{-}oral\,4\text{-}R}$	.71 ± .03	.78	.55 ± .04	.61
$r_{visual\text{-}oral\,5\text{-}R}$	.72 ± .03	.84	.71 ± .03	.81

We find for both corrected and uncorrected scores the oral 5-*R* correlates highest with the visual. The oral 4-*R* seems superior to the oral 3-*R*. Again, the size of the probable errors reduces the significance of the findings, but the evidence (if correlation with the visual test can be taken as evidence) seems to rather conclusively indicate a relative lack of validity on the part of the oral three-response, either corrected or uncorrected, and for the oral 4-*R*, particularly when uncorrected.

It will be recalled that the time order of presenting the tests was: Visual 5-*R*, oral 3-*R*, oral 4-*R*, and oral 5-*R*. One might conclude that this time order was such as to favor the five-response oral test, since the subjects had had considerable practice with the oral procedure by the time they took it. In the attempt to determine whether there was such practice effect we have made a study of the *wrongs* and *omissions* found in the early part of each test. Presumably if the

[1] Ruch, G. M.: "The Improvement of the Written Examination." Scott, Foresman, and Co., 1924, Chap. V, pp. 106–114.

directions were misunderstood, if the subjects did not grasp the procedure, we would expect to find this show up in the early part of the test, either in the form of omissions (blanks left) or wrong response words written in. Further, it seems safe to assume that the number of such omissions or wrongs found in the early part of the test (where the words are relatively easy) would be in some degree a measure of the misunderstanding. That is, if there are more wrongs or omissions in the first part of the oral 3-*R*, which was the first oral test administered, than in the first part of the oral 5-*R*, which was the last to be administered, we may assume a practice effect; and, likewise, if there are no differences in the number of omissions in the early part of the several oral tests we may assume understanding from the beginning and the order of presenting the tests is not coloring our findings.

In Table IV the average number of omissions and the average number of wrongs for the first twenty-five items of each of the tests is presented.

TABLE IV.—SHOWING AVERAGE NUMBER OF OMISSIONS AND WRONGS FOR FIRST TWENTY-FIVE ITEMS IN EACH OF FOUR FORMS

	Average omissions	Average wrongs
Visual 5-*R*	.23	3.11
Oral 3-*R*	.26	2.37
Oral 4-*R*	.56	3.92
Oral 5-*R*	.38	4.74

Examination of the table indicates that the oral 3-*R* has a smaller number of omissions and a smaller number of wrongs than the other two orally presented forms, and it has fewer wrongs but slightly more omissions than the visually presented test. This is exactly what one would expect in the absence of practice effect, since the oral 3-*R* has been shown to be the easiest form. The small number of omissions and wrongs found in each case may be interpreted as indicating no appreciable practice effect due to the order of presenting the tests, and it seems that is might also be interpreted as further validation of the oral procedure, since the students seemingly grasp the procedure without difficulty.

As a result of this investigation the following tentative conclusions seem justified.

1. Multiple-response tests presented orally are but slightly more difficult than the same tests presented visually.

2. Multiple-response tests may be presented orally without seriously reducing the reliability.

3. Multiple-response tests presented orally tend to measure that which is measured by the same tests presented visually.

4. There is no improvement in the reliability of oral multiple-response tests when they are corrected for guessing but such correction seems to slightly increase the correlation with the visual test.

5. Four or five responses seem superior to three responses when the tests are presented orally.

6. Until further evidence is presented it seems wise to recommend for oral presentation the five-response test with no correction for guessing.

AUDITORIUM *VERSUS* CLASSROOM SHOWING OF MOTION PICTURES IN HISTORY TEACHING

DANIEL C. KNOWLTON AND J. WARREN TILTON

Yale University

The experiment to be described is one of the Visual Education experiments conducted by the Department of Education of Yale University. In many respects, such as the place and use of the pictures in the teaching unit, the tests used, and the plan of testing for gain and retention, the experiment is a repetition of one already described in detail.[1] Consequently, the experiment to be described in this article will be presented with a minimum of detail.

Eleven of the fifteen Chronicles of America Photoplays were incorporated into the history teaching of the seventh grade of the Troup Junior High School of New Haven during the school year 1928–1929. The experiment may be thought of as six separate experiments, for an objective test was made by D. C. Knowlton for each of six units of instruction and was administered immediately before and after each unit. Each test was designed to measure fully and in a worth-while way, the entire unit of teaching in which the test was used. The tests were validated against expert opinion. In Table I are listed for the six units, the nature of the teaching unit, the month in which the teaching was done, the number of lessons or teaching periods included between the initial and final test, and the photoplays used. All tests were given for a third time in September, 1929. Results will be reported for the six units separately and combined.

The experiment was designed to measure the relative effectiveness of classroom and auditorium showing. To this end an effort was made to hold all factors constant except the experimental factor. The experiment was carefully supervised throughout by Mrs. Janet Marshall who, as an observer in the earlier study referred to, was thoroughly familiar with the plan and the nature of the experimental control needed. School cooperation was hearty and intelligent and experimental conditions were good.

At the request of the writers, Dr. William H. Martin, principal of the school, had so sectioned the grade and assigned classes to teachers, that each of three teachers taught two classes of the same

[1] Knowlton, D. C. and J. W. Tilton: "Motion Pictures in History Teaching." New Haven, Conn., Yale University Press, 1929, Part I.

TABLE I.—DESCRIPTION OF THE SIX PARTS OF THE EXPERIMENT

Teaching unit	Nature of the unit	Month in which teaching was done	Number of lessons between initial and final test	Photoplays used
I	Discovery	Oct.–Nov.	5	Columbus
II	Settlement	Dec.	6	The Pilgrims The Puritans
III	Comparison, European and Continental life	Jan.	8	Jamestown
IV	Extention of English domain	Feb.	3	Peter Stuyvesant
V	French and Indian War	Mar.	6	The Gateway to the West Wolfe and Montcalm
VI	Revolution	April May June	21	The Eve of the Revolution The Declaration of Independence Yorktown Vincennes

ability, matched pupil for pupil. The measure of ability used was educational age obtained by the use of the New Haven School Performance Test, Form 4. This test is a battery of nine tests similar to the Stanford Achievement Test. Sabina H. Connolly[1] found a correlation of $+.85 \pm .08$ between scores on the New Haven test and scores on the Terman Test of Mental Ability.

No measures of gain nor of retention were included for a pupil unless corresponding measures were available for the other pupil of the matched pair. No data for a pair of matched pupils was included unless those pupils were present for the retention test in September following the experiment. Finally, when proportionality of sexes had been equated, data were available for seventy-two matched pairs. The equality of these two groups of seventy-two pupils each is shown in Table II.

Educational quotients, initial history test scores, and school marks earned in English and mathematics during the year of the

[1] Connolly, Sabina H.: "The Reading Vocabulary of High School Students." Unpublished Masters Essay, Yale University.

experiment were compared as a check upon the equality of the groups as determined by educational ages. School marks were averaged by calling $A = 5$, $B = 4$, $C = 3$, $D = 2$, and $F = 1$. It should be noted that the groups were of above average ability as shown by their average educational quotients.

TABLE II.—EQUALITY OF GROUPS

	Auditorium group	Classroom group
Number of pupils	72	72
Number of boys	37	36
Number of girls	35	36
Mean educational age in months	157.4	156.9
SD educational age in months	10.1	10.0
Mean educational quotient	115.1	115.3
Mean initial score, Test I	17.4	15.9
Mean initial score, Test II	27.9	29.0
Mean initial score, Test III	37.0	36.0
Mean initial score, Test IV	31.3	31.6
Mean initial score, Test V	24.8	25.4
Mean initial score, Test VI	28.1	30.0
Mean English mark	3.4	3.5
Mean mathematics mark	3.7	3.6

The seventy-two pupils in the auditorium group saw the photoplays at the same time. In fact, with the three sections of which they were a part there were in the auditorium at the same time three other sections of the seventh grade. This arrangement was made to produce a contrast of numbers between the "classroom" method and the "auditorium" method. By the former method a thirty-five-pupil section saw the photoplays informally under the usual classroom conditions, in the room in which they regularly met for the history period. By the latter, about two hundred ten pupils more formally saw the photoplays in the auditorium. Many questions were asked in the classroom, none in the auditorium.

In the classroom, the machine was an Eastman Kodascope or a DeVry Portable, the film was sixteen millimeters in width, the screen was a TransLux or Daylo, and the picture was about sixteen by twenty-four inches. In the auditorium, two Powers A.C. machines were used in the booth, the film was standard width, a silver screen was used, and the picture was about nine by twelve feet.

In the classroom, the teacher used the part of the period preceding or following the showing and in between reels as she saw fit. Usually this time, amounting to about ten minutes, was spent in answering and discussing pupils' questions. In the auditorium there were no intervals between the reels, the alternate use of two projectors making continuous showing possible. The parts of the period preceding or following the showing of the picture were used as described in the following paragraphs.

The showing of *Columbus* and *Jamestown*, both four reel pictures, required the whole period.

The showing of *The Pilgrims* was preceded by a short talk by D. C. Knowlton in which he told the time covered by the photoplay and stressed the fact that the picture centered about six things—a humble cottage, a tiny ship, an unfriendly land, a trying winter, some old friends, and belief in God. These points were noted on a blackboard. After the picture was shown, three boys came to the platform with a rope eighty-six feet long, and stretched it out to show the length of the Mayflower. Then one hundred pupils came to the platform to show the numbers of passengers.

The Puritans was also preceded by a talk by D. C. Knowlton. He began by drawing a key on the blackboard as an indication of what the picture would do in serving to make the pupils understand the Puritans. He told why the Puritans were important and why they were more difficult to understand than the Pilgrims. He drew a small circle to represent Boston and the country immediately surrounding it, and a large circle to represent England, and then drew arrows connecting the two to show how events in one place acted upon events in the other. He ended by asking the six sections present to think up a new title for the picture and to submit their titles to Dr. Martin, the principal, who would select the best.

Peter Stuyvesant was preceded by a very short program. The stage was decorated with Dutch figures and scenes. A boy came to the platform and said, "By 1664 the English had settled the Atlantic seaboard. Here the English language was used and English customs followed. In the heart of this English territory was a thriving Dutch community where Dutch customs were followed, the Dutch language spoken, and boys and girls danced Dutch dances. Two members of our class will now dance a Dutch dance." A girl and boy in Dutch costume danced, and a poem *The Wooden Shoe* was recited.

The Gateway to the West was preceded by a blackboard talk by one of the teachers. She explained the meaning of Gateway in connection with the picture and briefly recalled the French explorations and settlements, drawing the St. Lawrence, the Great Lakes, the Mississippi, Ohio, Alleghany, and Monongahela Rivers, and the Appalachian Mountains. She drew a circle around the junction of the Alleghany and Monongahela and explained its importance as the Gateway. She ended by saying, "Now here is the question; Which nation will get to this gateway first, the English moving west, or the French moving down the rivers? Watch and see which one gets there first and what happens."

Wolfe and Montcalm was preceded by a short program arranged by another teacher. Ten pupils carried placards labeled respectively Vergor, Vaudreuil, Montcalm, French, Leader, English, Wolfe, Pitt, Anson, Saunders. They separated into French and English groups and each character announced himself in a short sentence. The spokesman then said, "When William Pitt became Prime Minister of England he declared he would carve out an empire for England. For this he would need a sword. Lord Anson should be the hilt of the sword, Saunders the blade, and the Wolfe the point." The characters named arranged themselves about a paper sword, Anson holding the handle and Wolfe the point. The spokesman concluded, "Watch the picture carefully and see how the sword moves."

The Eve of the Revolution was preceded by a short program arranged jointly by the same two teachers. Two groups of children came to the platform. Each pupil in the first group carried a placard. Their leader stepped forward and said, "These people represent some important events in the Revolution." The children, in turn, announced that they represented The Stamp Act, Boston Massacre, Boston Tea Party, First Continental Congress, Declaratory Act, Navigation Acts, Townshend Acts, Committees of Correspondence, and Quartering Act: The second group, eight girls, recited Longfellow's *Paul Revere's Ride*, each girl reciting a part.

The Declaration of Independence was followed, but not preceded, by a program. Immediately at the close of the picture a child went to the platform and said, "The Second Continental Congress will now meet." A second boy stepped forward and said, "I am John Hancock, President of the Second Continental Congress. As I call the names, the delegates from the different states will step forward." Thereupon he called the names of the thirteen colonies. One child in each delega-

tion had a placard giving the name of the state. The child with the placard in each case remained at the front of the platform but the others simply stepped forward to show the number of representatives from that particular state. Each child with a placard mentioned an item of interest connected with his state. For instance, the child representing Pennsylvania noted the fact that Pennsylvania had the largest delegation and that one of its members was Benjamin Franklin. When all thirteen states had been called, the children left the platform. A boy came to the front of the platform and said, "We will now have the reasons for the Declaration of Independence. Twenty-seven children came to the platform and stood in a row. In turn they all stepped forward saying, "I am reason 1," "I am reason 2," etc. Reasons 11, 12, 14, 17, 22 and 25 after giving their number said, "You saw me in the picture." After the twenty-seventh child had stepped forward they said in unison, "If you want to know what these reasons are, look in Gordy's history."

Yorktown was preceded by a brief program arranged again by the same teachers. Seven children went to the platform each representing one year of the Revolution. Each described briefly the important

TABLE III.—COMPARISON OF GAINS MADE BY THE AUDITORIUM AND CLASSROOM GROUPS

	Auditorium group			Classroom group			Classroom gain as percentage of auditorium gain	Difference in gain	SD[1] of difference
	Mean gain	Number of gains, averaged	SD of gains	Mean gain	Number of gains, averaged	SD of gains			
Teacher I, all units....	15.7	111	10.1	17.5	111	10.7	111	1.8	1.4
Teacher II, all units...	13.7	144	9.6	17.0	144	9.6	124	3.3	1.1
Teacher III, all units..	16.8	101	10.3	22.1	101	13.0	131	5.3	1.3
Unit I, all teachers....	8.0	66	4.5	10.0	66	4.1	125	2.0	.7
Unit II, all teachers...	14.9	57	7.4	16.7	57	5.9	112	1.8	1.3
Unit III, all teachers..	22.9	59	11.7	27.7	59	13.1	121	4.8	2.3
Unit IV, all teachers...	11.3	61	8.2	12.3	61	7.4	108	1.0	1.4
Unit V, all teachers....	12.7	56	6.2	14.4	56	6.1	113	1.7	1.2
Unit VI, all teachers...	22.3	57	10.4	26.3	57	12.3	117	4.0	2.1
All teachers, all units..	15.2	356	10.9	17.7	356	11.1	116	2.5	.8

[1] SD of the difference for this table and for Table IV was computed as equal to $\sqrt{(SDM_{NI})^2+(SDM_{NII})^2}$

events of their year. Seven other pupils stepped forward, each bearing a placard labeled respectively General George Washington, Count de Grasse, Rochambeau, Lord Cornwallis, Sir Henry Clinton, John Jouette, and Colonel Tarleton. Each described himself briefly.

Vincennes was shown before the program. After the picture eight pupils went to the platform three standing to the left and five to the right of the blackboard which occupied the center of the stage. The first pupil at the left was Patrick Henry and the second George Rogers Clark. Each pupil announced himself. The third pupil on the left went to the blackboard and traced the route followed by Clark. The five pupils at the right represented the states which were made from the Northwest Territory. Each carried a large pasteboard outline of the state he represented and each announced the name of his state—Ohio, Wisconsin, Illinois, Indiana, and Michigan respectively.

The gains made under the two methods are shown in Table III.

TABLE IV.—COMPARISON OF RETAINED GAINS MADE BY THE AUDITORIUM AND CLASSROOM GROUPS

(Retained Gain = Retest Score, Made in September of the Following School Year, Minus Initial Score)

	Auditorium group			Classroom group			Classroom gain as percentage of auditorium gain	Difference in gain	SD of difference
	Mean gain	Number of gains, averaged	SD of gains	Mean gain	Number of gains, averaged	SD of gains			
Teacher I, all units....	7.9	111	7.6	8.9	111	8.9	113	1.0	1.1
Teacher II, all units...	7.4	144	9.4	9.5	144	9.0	128	2.1	1.4
Teacher III, all units..	6.6	101	9.6	9.2	101	9.2	140	2.6	1.3
Unit I, all teachers....	4.4	66	4.5	5.2	66	5.0	118	.8	8
Unit II, all teachers...	10.7	57	6.3	13.0	57	7.4	121	2.3	1.3
Unit III, all teachers..	8.6	59	10.6	13.7	59	11.1	159	5.1	1.8
Unit IV, all teachers...	2.7	61	7.2	4.7	61	7.3	174	2.0	1.3
Unit V, all teachers....	4.7	56	7.2	5.8	56	6.0	123	1.1	1.3
Unit VI, all teachers...	13.1	57	10.2	13.6	57	10.3	103	.5	1.9
All teachers, all units..	7.3	356	9.3	9.2	356	9.3	127	1.9	.7

Whether the data are considered separately for the three teachers or separately for the six units of instruction, there is no evidence for the superiority of the auditorium showing. For all teachers and all

units combined the gain made by the classroom group is sixteen per cent in excess of that made by the auditorium group, and the difference is statistically significant.

The facts for retention are shown in Table IV. Again the findings are consistently in favor of classroom showing. In fact, for each teacher, for four of the six units, and for all data combined, classroom showing was relatively more effective in producing increases in score which were still in evidence at the beginning of the following school year than it was in producing immediate gains.

No data are available from this experiment to serve as a basis for an explanation of the superiority found for classroom showing. It is the opinion of the writers, that the difference is due, not to the differences in physics involved, but to the differences in pupil attitude and activity which differentiate classroom and auditorium periods.

THE VALIDITY OF PERSONALITY RATINGS OF PRESCHOOL CHILDREN

HERBERT S. CONRAD

Institute of Child Welfare, University of California

Can a comprehensive knowledge of a preschool child be obtained from ratings made by presumably competent judges? To throw light on this problem, three nursery school teachers rated thirty children in a Behavior Inventory for Nursery School Children devised by the writer.[1] The Behavior Inventory consists of two hundred thirty-one trait-descriptions, a sample of which is given below. It will be noticed that some of the descriptions refer to traits which are less clear-cut and less readily observed than such a trait as, shall we say, "vigor of physical activity." In this respect, the sample traits are probably representative of a majority of those in the Inventory; and this, in turn, of the greater portion of children's actual behavior. Children's behavior rarely permits clear-cut, ready inference as to personality. *Rating scales which select the few traits permitting easy observation and interpretation must either remain frankly incomplete, or assume to achieve a degree of reliability and validity which is only too obviously unrepresentative.*

The directions to the three judges were to rate each child in comparison with others of his own chronological age.[2] The ratings were, of course, to be made independently, without conference or consultation. For each trait, a child was to be rated:

4, if *average* for his age.
3 or 5, if *differing slightly from the average for his age.*
2 or 6, if *distinctly exceptional.*
1 or 7, if *extreme and outstanding.*

Thus, for "vigor of activity" (which is one of the sample-traits described above), the rating 1 would denote an extremely energetic and vigorous child; the rating 7 would denote an extremely unenergetic child; 4 would denote the average.

[1] Acknowledgment is gratefully made for assistance from Dr. Harold E. Jones, Director of Research at the Institute of Child Welfare of the University of California; and from the three nursery school teachers of the Institute, Miss L. Allen, Mrs. G. N. Ludwig, and Miss E. Lewis.

[2] This instruction was satisfactorily carried out, as evidenced by the negligible correlation between ratings and children's chronological age. (Range of *CA*, 23–57 months; mean *CA*, 41.2 months; median, 41.5 months.)

SAMPLE TRAITS FROM THE BEHAVIOR INVENTORY

Vigor of Physical Activity.

Forceful, energetic, vigorous; runs fast, saws wood energetically, etc.; high energy output.			Average degree of vigor; average energy-level.			Activity is carried on at a low energy-level; play and work is of a mild, placid type.
1	2	3	4	5	6	7

Reaction to Social Pressure.

Accedes immediately to the demands (explicit or implicit) of the group.			Average. Responds with moderate promptness to reasonable demands of the group; but resists or becomes indifferent to the group's demands if it expects him to serve as a butt of amusement, or perform disagreeable tasks, etc.			Indifferent to the demands and expectations of the group. Individualistic and independent.
1	2	3	4	5	6	7

Standards in Work and Play.

Child evidently has exceptionally high standards; tries to achieve a high standard of perfection in his work and play; tries to be thorough, and achieve polish and finish. Likely to sacrifice speed for quality. Not at all sloppy, careless, rough-and-ready, uncritical, etc.			Average. Child's standards in work and play are not higher (nor lower) than those of others of his age. Child sacrifices thoroughness or polish when in a hurry.			Child has exceptionally low standards in his work and play; does not try to be thorough; neglects polish and finish in his activities; accepts low quality performance; is sloppy, or careless, or rough-and-ready, or uncritical, hasty, etc.
1	2	3	4	5	6	7

Use of Facial Expression; Effectiveness.

Child's facial expressions are clear, readily understood, exceptionally effective in communicating feelings, wishes, invitations, etc.			Average. Facial expressions are clear and effective for major emotions, but less clear in communicating feelings, wishes, invitations, etc.			Child's facial expression is clear only in the case of the gross, simple, major emotions. Facial expression (or lack of facial expression) is exceptionally ineffective or misleading in communicating shades of emotion, or wishes, invitations.
1	2	3	4	5	6	7

It was suggested to the teachers that an appropriate distribution of their ratings would be as follows: about 2 per cent of the ratings should be 1; about 8 per cent, 2; about 23 per cent, 3; about 33 per cent, 4; about 23 per cent, 5; about 8 per cent, 6; and about 2 per cent, 7. In order to avoid halo effects so far as possible, all the children were to be rated on one trait at a time, rather than a single child on all traits at a time.

One other direction requires notice: For each child being rated, the teachers were instructed to *star* their ratings of any traits which they

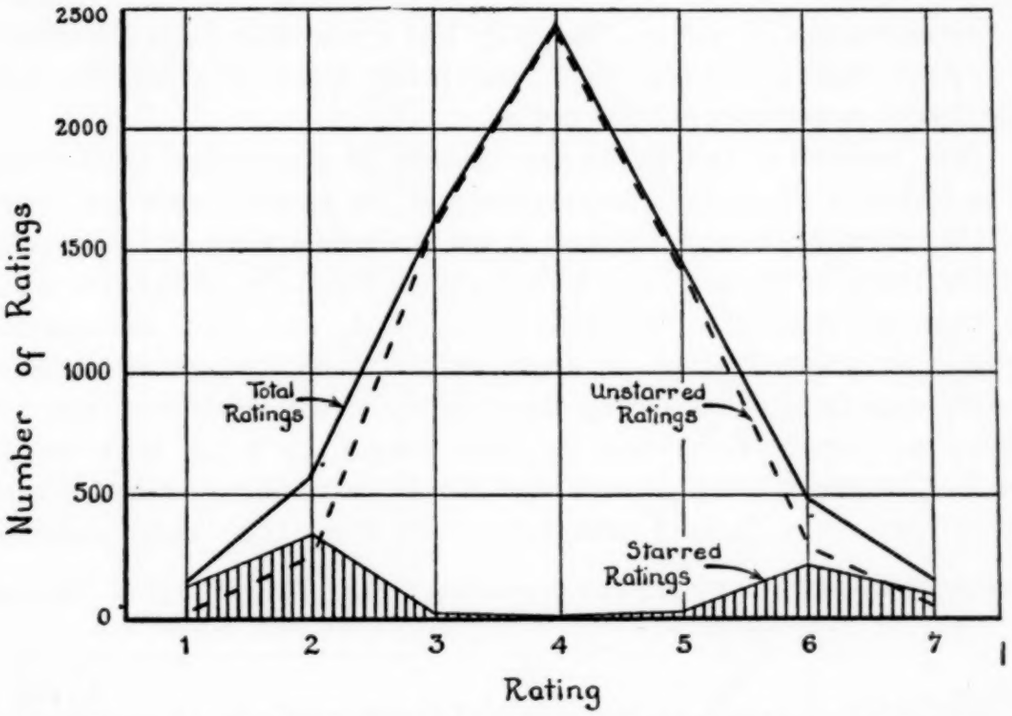

FIG. 1.—Frequency distributions of judge *C*'s starred, unstarred, and total ratings of thirty nursery school children.

regarded of "central or dominating importance in the child's personality"—such that they would "give prominence to these traits in a description of the child's personality," or would "consider that these traits contribute in an outstanding way to the success or failure of the child's adjustment." Thus, if judge *A* regarded a certain child's vigor of activity to be of sufficient importance in the child's personality, she would star her rating of this trait for that particular child—but not necessarily for any other child. Any trait-rating could be starred, depending only on the teacher's judgment of its importance in the personality of the child in question. Starred trait-ratings are hereafter sometimes referred to as "critical" ratings, unstarred trait-

ratings as "non-critical." Of judge *A*'s total ratings, twenty-one per cent were starred; of judge *B*'s, nine per cent; and of judge *C*'s, eleven per cent. A graph of *C*'s starred, unstarred, and total ratings is given in Fig. 1. These distributions are representative of those from the other judges, except that *C* starred fewer trait-ratings than did *A*. A mathematical fit has indicated that the total ratings follow fairly closely the form of the normal curve.

The greatest interest, of course, attaches to the reliability and validity of the critical (starred) trait-ratings. These are the ratings of the significant traits for the children in question. Too often, in previous studies of ratings, the judge has apparently been expected to rate each child accurately on a great many traits of which the judge justifiably never took special notice.

One method of indicating the validity of the critical trait-ratings is to correlate the starred trait-ratings of one teacher with the ratings by the other teachers. Thus, if judge *A* made a total of 1448 starred trait-ratings for the total group of thirty children, the correlation would be between these starred trait-ratings of *A*, and the corresponding ratings (whether starred or unstarred) of judge *B*, or of judge *C*. Another method of indicating the validity of the critical ratings is to find the average difference (without regard to sign) between one teacher's starred trait-ratings, and the corresponding ratings by the other teachers. Table I presents results from both these methods.

TABLE I.—AGREEMENT BETWEEN JUDGES, FOR TRAIT-RATINGS STARRED (REGARDED AS CRITICAL) BY AT LEAST ONE JUDGE

Trait-ratings starred by judge	No. of trait-ratings starred[1]	Correlation between judges	r	Mean diff. between judges' ratings
A	1448	$A \times B$	.72	1.26
A	1448	$A \times C$	.67	1.31
B	617	$B \times A$	.82	.88
B	615	$B \times C$	.78	.98
C	760	$C \times A$	.69	1.25
C	768	$C \times B$	.72	1.27

[1] The slight discrepancies in numbers of cases arise from occasional omissions of trait-ratings by a judge.

A given trait is more assuredly of critical importance in a given child's personality, if at least two, or all three, of the nursery school teachers have starred their judgments of this trait for the child in

question. In the clinical study of a child, such doubly- or triply-starred ratings would of course receive the greatest consideration. The validity of the doubly- and of the triply-starred ratings is readily indicated by finding the correlation (or the average difference) between the ratings of one judge, and those of the others. Tables II and III present the results.

TABLE II.—AGREEMENT BETWEEN JUDGES, FOR TRAIT-RATINGS STARRED (REGARDED AS CRITICAL) BY AT LEAST TWO JUDGES

Correlation between ratings of judges	No. of trait-ratings	r	Mean diff. between judges' ratings
$A \times B$	589	.85	.84
$A \times C$	589	.81	.94
$B \times C$	589	.79	.98

TABLE III.—AGREEMENT BETWEEN JUDGES, FOR TRAIT-RATINGS STARRED (REGARDED AS CRITICAL) BY THREE JUDGES

Correlation between ratings of judges	No. of trait-ratings	r	Mean diff. between judges' ratings
$A \times B$	146	.95	.47
$A \times C$	146	.93	.60
$B \times C$	146	.96	.42

The figures in Tables I–III appear rather gratifying. Evidently, fairly high correlations between judges' ratings, even up to .96, may be obtained for traits which are significant in the individual child's personality. Pooling the ratings of the three judges would of course make the correlation coefficients still higher.

Of possibly greater importance, to our mind, than any purely statistical check which can be applied, is the practical, clinical test by which the value of the starred trait-ratings has been ascertained. Every term, the nursery school teachers of the Institute of Child Welfare prepare a personality study of each child. This work has not only been considerably expedited, but also improved by the compilation, for each child, of the ratings which were starred by (at least) one of the three judges.

A point of major significance in the present study is *the dependence of the reliability and validity of a rating scale upon the individual children being rated.* It appears that most of the nursery school children can

be rated fairly reliably and validly; but a few, even though rated by the same judges and on the same scale, are judged with little reliability, and practically no validity. The agreement between judges *A* and *B* in their ratings of the individual children is given in Table IV. Table IV is based on the total (both starred and unstarred) ratings of each judge on the entire Inventory of 231 traits.[1] The figures in Table IV are typical of the correlations between the other pairs of judges.

TABLE IV.—AGREEMENT BETWEEN TWO JUDGES IN THEIR TOTAL RATINGS OF INDIVIDUAL CHILDREN

Child No.	*r* between judges' ratings	Av. diff. between judges' ratings	Child No.	*r* between judges' ratings	Av. diff. between judges' ratings
1	2	3	1	2	3
1	.44	.73	16	.43	.88
2	.42	.69	17	.54	.66
3	.34	.75	18	.41	.64
4	.39	.76	19	.50	.87
5	.51	.92	20	.33	.91
6	.71	.86	21	.52	1.26
7	.14	.79	22	.56	.96
8	.71	.87	23	.29	1.13
9	.51	.97	24	.43	.71
10	.54	.98	25	.23	.60
11	.72	.77	26	.47	.87
12	.48	.87	27	.32	.95
13	.58	.78	28	.49	.80
14	.48	.95	29	.38	.95
15	.78	.90	30	.50	1.03
Mean *r*	.47		Mean av. diff	.86	
Median *r*	.48		Median av. diff	.87	

It can be seen that the coefficients run from .14 to .78, with a median of .48.[2] If the ratings by the three teachers were pooled, then for

[1] We are using the total ratings of each judge for each child, rather than the critical ratings only, in order to insure that the correlation between the judges for each child will be fairly reliable. The starred ratings for each child are not, as a rule, sufficiently numerous to yield reliable correlations between judges for any individual child.

[2] The median correlation between judges *A* and *C* is .45; between *B* and *C*, .47. If the ratings for all the children are combined (as in common practice they

all the traits (both starred and unstarred, for the child being rated), the average agreement between the ratings by one set of three judges and another may be estimated as over .70.[1]

Our figures on the reliability of the ratings for each child are unfortunately not so adequate as those which have just been presented for validity. Only thirty-one traits of the Behavior Inventory were re-rated (after an interval of about one to two months) by the three teachers.[2] The average initial rating for each child was determined

TABLE V.—RELIABILITY OF AVERAGE OF THREE TEACHERS' RATINGS OF EACH CHILD (EACH CHILD RATED AND RE-RATED IN THIRTY-ONE TRAITS)

Child No.	*r* between average rating and average re-rating	Av. diff. between average rating and average re-rating	Child No.	*r* between average rating and average re-rating	Av. diff. between average rating and average re-rating
1	2	3	1	2	3
1	.80	.42	16	.76	.40
2	.75	.38	17	.60	.46
3	.83	.34	18	.83	.27
4	.84	.32	19	.76	.52
5	.79	.46	20	.57	.43
6	.94	.45	21	.81	.61
7	.32	.31	22	.84	.45
8	.92	.44	23	.74	.41
9	.82	.44	24	.77	.41
10	.92	.33	25	.42	.28
11	.95	.33	26	.83	.48
12	.83	.52	27	.70	.51
13	.74	.44	28	.90	.41
14	.76	.63	29	.74	.34
15	.94	.48	30	.92	.42
Mean *r*	.78		Mean av. diff		.42
Median *r*	.81		Median av. diff		.43

generally are) into a single scatter-diagram, the correlations between *A* and *B*, *A* and *C*, and *B* and *C* are .54, .47, and .50 respectively. Shepard's correction has not been applied.

[1] Kelley, T. L.: "Statistical Method." Formula 154. Macmillan, New York, 1924.

[2] These thirty-one traits may, by a number of criteria, be taken as reasonably representative of the total Behavior Inventory.

for each of the thirty-one traits, and these averages correlated with the average re-rating. This gives us a reliability coefficient of the ratings (both starred and unstarred) of each child, based on thirty-one traits. Table V presents these reliability coefficients: it can be seen that they range from .32 to .95. The mean difference between the average initial-rating for each child and the average re-rating, is also presented in Table V (Column 3).

A point of interest to the clinical psychologist is the discrepancy between the results as stated in terms of correlation coefficients and in terms of average differences. This discrepancy is quite pronounced. Thus, the correlation between the coefficients of Column 2, Table IV, and the average differences in the column adjoining, is only −.18; between Columns 2 and 3 of Table V, the correlation is −.02. It must be remembered that the correlation coefficient is a relative, the average difference is an absolute, measure. The correlation coefficient is based on the difference between judges' ratings, in relation to the standard deviations of these ratings; in this sense, the coefficient of correlation makes allowance for such agreement as might have occurred by chance. The average difference, on the other hand, is simply the absolute average difference. Which of these is the appropriate measure for clinical use? This would appear to be a debatable question; doubtless one measure supplements the other, so that both are useful. From the point of view of the theory of measurement, the correlation coefficient appears to us, on the whole, the more appropriate measure. It is principally in terms of correlation that the present paper has been written.

A final word on the absence of experimental or observational verification of the validity of the teachers' ratings. Such verification would be desirable, if it could be validly and economically secured. *Relatively few* experimental situations, however, correspond to the genuine situations of life—and we see little point in judging the validity of ratings of two hundred thirty-one traits, on the basis of experimental results from a restricted, selected few. An additional difficulty is the problem of securing reliable scores from experimental tests of behavior. Observational verification, while ordinarily more valid than experimental, is perhaps even more expensive. Observational verification of the teachers' ratings of certain traits is now in progress.

Lest we be considered excessively enthusiastic in our support of ratings of character, we wish to add that the preschool child in the nursery school may be a relatively easy subject to rate. The nursery

school child's reactions are comparatively naive and unconcealed; he is given plenty of opportunity in the school to display a large portion of his character; the child has not yet learned the bewildering number of extremely specific reactions and adjustments which so complicate the behavior of an adult; and finally, the nursery school teachers are always about, ready to observe the child's confession of behavior, and unencumbered by administrative requirements which tend (in the higher school-grades) to focus attention upon curricular routine rather than on the child's personality. In view of these circumstances, it may be difficult to duplicate in a sample of older children or of adults the results of the present study. This does not alter the fact, however, that at least for the great majority of nursery school children, the answer to the question, "Is the rating of human character practicable?" must be an emphatic affirmative. The way appears definitely open, therefore, to begin to test the widely entertained hypothesis concerning the crucial importance, for future character, of the preschool years.

SUMMARY

Ratings of thirty children (aged twenty-three to fifty-seven months) were made by three nursery school teachers on a Behavior Inventory for nursery school children consisting of two hundred thirty-one trait-descriptions. Each of the two hundred thirty-one traits was rated on a seven-point scale. A novel feature of the ratings consisted in each teacher's starring her ratings of any traits which she considered of central importance for the particular child being rated. The greatest interest attaches to the reliability and validity of these starred trait-ratings, for these are the ratings of the traits which are regarded as significant for the child in question. The agreement among the teachers concerning significant traits is satisfactory—varying for traits held to be significant by all three judges, from $r = .93$ to $r = .96$.

The reliability and validity of the Behavior Inventory varies importantly from child to child. This dependence of the worth of a rating scale upon the *individual children* being rated, has not hitherto received adequate emphasis.

Many of the trait-descriptions of the Behavior Inventory refer to traits which are less clear-cut and less easily observed than the characteristics conventionally included in rating scales. This is a consequence of the relative completeness of the Inventory. Traits difficult to judge have not been excluded simply because of their difficulty.

The favorable results obtained from the Behavior Inventory, despite its unselected character, have persuaded us that the Inventory represents a valuable method of attack upon many of the problems of genetic psychology. Of special interest is the possibility of a scientific and systematic determination of the influence of the pre-school years upon later character development.

SCORING OBJECTIVE TESTS*

NOEL B. CUFF

Eastern Kentucky State Teachers College

1. INTRODUCTION

The protagonists of new-type tests have made numerous claims relative to the reliability and to the validity of objective tests, and concerning the ease of administering and of scoring such tests. Ruch[6] says the increased ease of scoring makes it possible for the scoring to be done by clerical workers without previous experience or training, or, in some cases, by pupils. He states too that objective tests of forty to seventy-five items can be scored at rates of twenty to thirty or more per hour. Lang[3] says the objective test eliminates most of the drudgery that teachers ordinarily endure in scoring essay-examination papers. Greene and Jorgenson[2] say: "A perfectly objective test may be scored repeatedly by one person with practically no variation in the scores, or it may be scored by a large number of persons with practically no variation in scores assigned."

From such quotations, it would seem one is justified in concluding that objective tests can be scored (1) rapidly and (2) accurately. But we find some authorities who say it is sometimes even difficult to score true-false tests, that care should be taken to facilitate scoring, and the like. A survey of the directions for scoring standard tests shows that accurate scoring of objective tests is at least considered difficult. These directions are given in the manual of the "Otis Self-Administering Tests of Mental Ability": "Do not trust the counting of correct answers only, as it is very easy to make a mistake. The checking of correct answers should be gone over by a second scorer, for even the best scorers will make mistakes." The manual of directions for the "National Intelligence Tests" contains this: "Where accuracy is desired, all scoring should be checked by a second scorer. It has been found that intelligent and experienced scorers who thoroughly understand the rules average about one error to an examination paper. Inexperienced scorers make several times as many mistakes."

Pintner[5] has attempted to answer two questions: (1) "To what extent are our methods of scoring 'fool-proof'? and (2) How much

* Read at the 27th annual meeting of the Southern Society for Philosophy and Psychology, in joint session with Section I, Psychology of the American Association for the Advancement of Science, New Orleans, La., December 30, 1931.

emphasis must be given to the matter of scoring in the training of students in testing?" He had graduate students score two blanks of the "National Intelligence Tests" that had been filled with various difficulties. "An alarming range of scores" was given to the same paper, the range for Form 1 was three to eighty-five, but it was not an ordinary paper. The errors, Pintner states, were due to differences in interpretation, to oversight, and to mistakes in adding and subtracting.

Dearborn[1] found, by having five hundred thirty Dearborn Tests rescored, that seventy-three per cent of all tests contained error. He says: "Scoring is monotonous work, even if one be interested in it, and error is bound to creep into work whose salient feature is monotony."

In view of the preceding survey it is obvious that an attempt to determine the rate and accuracy of scoring objective tests by certain methods will be of interest.

2. PROBLEMS AND PROCEDURE

The present study, which is a preliminary report, is concerned with two questions: (1) How may objective tests be scored most rapidly? (2) How may objective tests be scored most accurately?

In an attempt to answer these questions a survey was made of pertinent suggestions relative to true-false and multiple-choice tests. Ruch[6] states that the device, of placing the numbers of the choices in a straight column, which was invented by Otis, represents the most rapidly scored multiple-choice technique ever developed. Paterson,[4] in discussing true-false questions, says: "The encircling of the letters 'T' or 'F' or the words 'True' or ' False' before the statements is perhaps the better method, and it is better to have the pupil encircle or underline these than it would be to require him to write the letters or words in a column, partly because such letters are easily confused in scoring and partly because it is easier to score a paper when the answers zigzag than it is when they are in a single column." Clapp and Young have an "Answer Booklet" that can be used with either true-false or multiple-choice tests and they insist it can be scored at least four or five times faster than any test that corresponds with it exactly.

After becoming acquainted with the Clapp-Young tests and with Form 17 of the O. C. A., Intelligence Tests, an Answer Sheet was devised that enables one to construct true-false or multiple-choice tests, and yet have the advantage of mimeograph marking.

To determine how rapidly and accurately true-false tests can be scored, two methods of recording responses were used. First, two

mimeographed tests, Forms 1 and 2, of one hundred items each, so nearly equal in difficulty that the quartiles are the same, were given to one hundred seventy students at regular class meetings with the instructions: "Encircle the 'T' before each true statement and the 'F' before each false statement." Second, two mimeographed tests, Forms 3 and 4, approximately equal item for item in difficulty to Forms 1 and 2, were given to the same students with Answer Sheets that had these directions:

"Indicate, opposite number 1, your answer to the first question of a test by marking X on the number that is the same as the number of the correct one of the suggested answers. Then answer other questions in the same way.

"For a true-false test, mark an X on 1 for true and an X on 2 for false, unless you are told to use 4 for true and 5 for false."

The Answer Sheet has four columns of numbers; each containing question numbers followed by numbers for five possible choices. For example:

1.	1	2	3	4	5	26.	1	2	3	4	5	51.	1	2	3	4	5	76.	1	2	3	4	5
2.	1	2	3	4	5	27.	1	2	3	4	5	52.	1	2	3	4	5	77.	1	2	3	4	5

A scoring key was made for Forms 1 and 2 of the test by encircling T's and F's on a copy of the test and then cutting off the margin. Such a key was used by each scorer.

The Answer Sheets for Forms 3 and 4 of the test had small circles mimeographed around the correct choices. If an X was on a number other than the one encircled, the response was incorrect; the circles contained X's when the answers were correct.

Multiple-choice tests of one hundred statements, with five possible answers for each statement, were also given to the same students. The technique devised by Otis for recording answers was used for about half of the tests—answer strips were used in scoring these—and the Answer Sheet was used—in the manner that has been described—for the others.

Three individuals scored about one hundred papers of each of the four types of responses—two types of true-false and two types of multiple-choice responses.

3. RESULTS

The results are presented in Tables I and II. Table I shows that Scorer *A* averaged scoring six Answer Sheets in the time which was

required for him to score one test with the T's and F's encircled before the items; the respective times per paper in seconds are thirty and one hundred ninety two. Scorer *C* used forty-three seconds per paper when the answers were placed on Answer Sheets and one hundred twenty-six seconds per paper when the T's and F's were encircled before the items. Scorer *A* scored accurately seventy papers out of one hundred when the Answer Sheets were used and only forty out of one hundred when

TABLE I.—RATE AND ACCURACY OF SCORING TRUE-FALSE TESTS

Scorer	Arrangement of answers	Scoring time per test (seconds)	Size and number of errors per 100 tests								
			0	1		2		3		4	
				+	−	+	−	+	−	+	−
A	1*	30	70	22	4	3	..	..	..	1	
	2†	192	40	30	10	20					
B	1	32	67	27	2	4					
	2	198	50	40	10						
C	1	43	87	5	8						
	2	122	40	40	10	7	..	3			

* Answer indicated by making an X on the Answer Sheet.
† Answer indicated by encircling T or F before the item.

the letters, T and F, were encircled. Of the tests that were inaccurately scored by *A*, twenty-two Answer Sheets were overscored one point, four were underscored one point, three were overscored two points, and one was overscored three points. The errors in scoring, however, were even more numerous when the answers were indicated

TABLE II.—RATE AND ACCURACY OF SCORING MULTIPLE-CHOICE TESTS

Scorer	Arrangement of answers	Scoring time per test (seconds)	Size and number of errors per 100 tests								
			0	1		2		3		4	
				+	−	+	−	+	−	+	−
A	1*	52	88	7	2	3					
	2†	231	71	21	6	2					
B	1	29	90	7	3						
	2	158	90	8	2						
C	1	72	90	7	1	2					
	2	165	83	11	3	3					

* Answer indicated by marking an X on the Answer Sheet.
† Answer indicated by writing the number of the choice before the statement.

on the tests. The Answer Sheets were scored more rapidly and more accurately by each of the three scorers.

Table II shows that when Answer Sheets were used for the multiple-choice tests the time was from less than one-half (Scorer *C* used seventy-two and one hundred sixty-five seconds for the two methods) to less than one-fifth (Scorer *B* used twenty-nine and one hundred fifty-eight seconds respectively) of the time required to score the tests when the answers were indicated by the Otis method. Scorers *A* and *C* made fewer errors in scoring the Answer Sheets, but the data show no marked differences for Scorer *B* in the number or the size of the errors for the two methods.

4. SUMMARY

The data presented in this study suggest these conclusions:

1. When the answers are indicated on an Answer Sheet so that they can be marked with a mimeograph, true-false and multiple-choice tests can be scored from two to six times more rapidly than when the answers are indicated opposite the test items by frequently recommended methods.

2. When answers are indicated on an Answer Sheet so that they can be marked with a mimeograph, true-false and multiple-choice tests can be scored more accurately, at a rapid rate, than when the answers are indicated by commonly used methods.

Since an Answer Sheet can be scored rapidly and accurately, it obviously has these advantages:

1. An Answer Sheet eliminates most of the labor of marking answers right or wrong.

2. Errors in marking the answers are completely done away with.

3. Since test blanks are not marked, they can be filed and used at a later date.

4. More reliable tests—more comprehensive and more accurately scored—can be given.

5. Tests can be scored promptly and pupils can be informed promptly of their mistakes.

6. One can have several patterns or stencils for true-false and for multiple-choice tests so that students cannot memorize the keys.

7. Since an Answer Sheet consists of only one page, it is inexpensive.

REFERENCES

1. Dearborn, W. F. and C. W. Smith: The Results of Rescoring Five Hundred Thirty Dearborn Tests. *Journal Educational Psychology*, Vol. XX, 1929, pp. 177–183.
2. Greene, H. A. and A. N. Jorgensen: "The Use and Interpretation of Educational Tests." 1929.
3. Lang, A. R.: "Modern Methods in Written Examinations." 1930.
4. Paterson, D. G.: "Preparation and Use of New-Type Examinations." 1927.
5. Pintner, R.: Accuracy in Scoring Group Intelligence Tests. *Journal Educational Psychology*, Vol. XVII, 1926, pp. 470–475.
6. Ruch, G. M.: "The Objective or New-Type Examination." 1929.

A STUDY OF HIGH SCHOOL GRADUATES WITH REFERENCE TO LEVEL OF INTELLIGENCE

CONSTANCE E. DOWD

Psychological Laboratory, Vocation Bureau, Cincinnati Public Schools

Group intelligence tests are now widely used for educational and vocational guidance. Experience with the use of tests in guidance work has pointed to the need of better information as to the implications of a given test rating in terms of subsequent educational outcomes. The present study was undertaken with a view to throwing light on the question of the predictive value of the group tests in connection with high school graduation.

HISTORICAL SUMMARY

Many studies of the relationship of intelligence test scores to school marks have been made, but relatively few investigations have been made concerning the relation between intelligence test scores and progress through school over a period of several years. Such studies as exist show a positive relation between intelligence level and pupil persistence in school. Probably the most frequently quoted study is that of Terman and Proctor.[1] In the light of a sixth and seventh year check-up in 1923 of one hundred seven pupils who entered high school in 1916–1917 and forty who entered eighth grade in 1917–1918 Proctor[2] states that it might have been predicted that one hundred per cent of the "very superior" (IQ one hundred twenty-five or over), ninety-six per cent of the "superior" (IQ one hundred fifteen to one hundred twenty-four), eighty-three per cent of the "above average" (IQ one hundred five to one hundred fourteen), seventy-five per cent of the "average" (IQ ninety-five to one hundred four), forty per cent of the "below average" (IQ eighty-five to ninety-four), and zero per cent of the "dull" (IQ below eighty-five) would complete a high-school course.

In 1917–1918 Proctor gave Army Examinations A and B and Army Alpha, Form 5, to two thousand high school pupils. Nine hundred fifty-five cases were followed up for two years. The median intelli-

[1] Terman, L. M.: "The Intelligence of School Children." Houghton Mifflin, Chap. VI, 1919, p. 90.

[2] Proctor, William Martin: "Educational and Vocational Guidance." Houghton Mifflin, 1925, p. 31.

gence quotient of the graduated group was one hundred twelve, of the group remaining in high school one hundred seven, and of the group that dropped out ninety-six. Of the one hundred fifty-three pupils who were graduated from high school four had intelligence quotients between eighty and ninety, twenty-two had intelligence quotients between ninety and one hundred, and one hundred twenty-seven had intelligence quotients of one hundred or over.

Pintner[1] states that there is evidence that selection in high school was and still is partly on the basis of intelligence, and that intelligence tests show that the high school pupil belongs in the upper half of the distribution of intelligence. He also notes that the more intelligent pupils tend to be found in the more academic courses and the less intelligent in the more practical courses, and that as our high schools are at present, children with intelligence quotients between ninety and one hundred will have great difficulty in graduating.

Dearborn[2] showed that the median intelligence quotient in three average New England towns rose from ninety-seven in Grade II to one hundred eight in Grade XII.

Book[3] tested 6188 seniors in three hundred twenty high schools in Indiana. Since he assigned arbitrary intelligence grades *A*, *B*, *C*, *D*, *E*, *F* it is difficult to compare his results with those where tests results are reported in terms of intelligence quotients or age percentile ranks.

GROUP STUDIED

Out of 4184 sixth grade pupils who were given group tests in 1923–1924 those who completed high school in 1929, 1930, and 1931 have been selected for study. As a basis of comparison for the group of high school graduates, Tables I and II show distributions of group test percentile ranks and chronological ages for the total group of 4184 sixth grade pupils.

In the high schools in and about Cincinnati, including public, private, parochial and night schools, seven hundred thirty-two of the original sixth grade group were graduated from a four-year high school course in 1930. In addition to this major group that completed high

---

[1] Pintner, Rudolph: "Intelligence Testing." Henry Holt, 1931, pp. 274–289.

[2] Dearborn, Walter Fenno: "Intelligence Tests." Houghton Mifflin, 1928, p. 155.

[3] Book, William F.: "The Intelligence of High School Seniors." MacMillan, 1922.

school at the normal rate without repeating or skipping a grade, a smaller group of one hundred forty-nine was graduated in 1931. This group had obviously repeated or missed a grade somewhere between the sixth grade and graduation from high school. A still smaller group of thirty-two pupils was graduated in 1929, having skipped one grade between the sixth grade and high school graduation.

It appears, therefore, that nine hundred thirteen pupils from the total number, 4184, tested in the sixth grade in 1923–1924 are known to have completed high school within at most seven years after the year of the sixth grade tests. In addition to this number twelve pupils are known to have been graduated from high schools not in Cincinnati. We cannot draw the negative inference that none of the remaining 3259 have completed, or will complete high school. However, our search of the records in and about Cincinnati has been sufficiently thorough so that we are convinced that we have in this number practically all members of the original group who completed a four-year high school course here within the period of seven years. We

TABLE I.—PERCENTILE RANKS ON OTIS ADVANCED EXAMINATION FOR 4184 SIXTH GRADE PUPILS, 1923–1924

PERCENTILE RANK	NUMBER OF PUPILS
95–100	108
90– 95	141
85– 90	180
80– 85	183
75– 80	193
70– 75	223
65– 70	209
60– 65	220
55– 60	229
50– 55	219
45– 50	225
40– 45	228
35– 40	228
30– 35	235
25– 30	234
20– 25	228
15– 20	252
10– 15	236
5– 10	254
0– 5	159
Total	4184
Median	45.87
Quartile deviation	23.22

have not included in the present study the twelve children graduated from high schools away from Cincinnati since we are unable to include all cases of this type from the original group of 4184.

TABLE II.—CHRONOLOGICAL AGES OF 4184 SIXTH GRADE PUPILS AT TIME OF ENTRANCE INTO SIXTH GRADE (SEPTEMBER, 1923)

CHRONOLOGICAL AGE	NUMBER OF PUPILS
17 yr. –17 yr. 6 mo	2
16 yr. 6 mo.–17 yr	4
16 yr. –16 yr. 6 mo	15
15 yr. 6 mo.–16 yr	39
15 yr. –15 yr. 6 mo	77
14 yr. 6 mo.–15 yr	145
14 yr. –14 yr. 6 mo	204
13 yr. 6 mo.–14 yr	277
13 yr. –13 yr. 6 mo	341
12 yr. 6 mo.–13 yr	458
12 yr. –12 yr. 6 mo	566
11 yr. 6 mo.–12 yr	736
11 yr. –11 yr. 6 mo	798
10 yr. 6 mo.–11 yr	379
10 yr. –10 yr. 6 mo	105
9 yr. 6 mo.–10 yr	32
9 yr. – 9 yr. 6 mo	4
8 yr. 6 mo.– 9 yr	2
Total	4184
Median	11 years 11.9 months
Quartile deviation	1 year 9.1 months

Ages as stated are correct within one-half month. Thus, the interval ten years, no months to ten years, six months is actually nine years, 11½ months to ten years, 5½ months. Medians are computed on the basis of the actual limiting values of the intervals.

DATA REGARDING INTELLIGENCE

The intelligence ratings used in the present study are percentile ranks and mental ages based on Otis' norms for the Otis Advanced Examination. The percentile ranks are determined in the light of the age of the child at the time of the test in the sixth grade.[1] As Otis has pointed out,[2] it is preferable to use percentile ranks rather than intelligence quotients for group test ratings because the scores on

[1] Otis, Arthur S.: "Manual of Directions for Advanced Examination." World Book Company, 1929, pp. 7 and 12.

[2] Otis, Arthur S.: "Manual of Directions for Otis Group Intelligence Scale." Edition 1929, p. 50.

group tests distribute themselves as widely at low age levels as at high age levels, so that a ratio such as an intelligence quotient is not an adequate measure.

The objections to using intelligence quotients on Otis Group Tests are also pointed out by Psyche Cattell[1,2] in two recent articles. If, however, one who is accustomed to thinking in terms of intelligence quotients wishes to compare the percentile ranks in the present discussion with Binet-Simon intelligence quotients, it is possible to refer to a table given by Otis in his Manual of Directions for the Advanced Examination.[3] This table is constructed on the basis of Terman's statement that the semi-interquartile range of the distribution of intelligence quotients is eight points. Thus a percentile rank of seventy-five would correspond to an intelligence quotient of one hundred eight, and a percentile rank of twenty-five to an intelligence quotient of ninety-two. A percentile rank of ninety would correspond to an intelligence quotient of one hundred fifteen, a percentile rank of eighty to an intelligence quotient of one hundred ten, a percentile rank of twenty to an intelligence quotient of ninety, and a percentile rank of ten to an intelligence quotient of eighty-four. This is only a rough approximation and, as has been pointed out above, it is preferable to think of the group test results in terms of percentile ranks for each age group.

The mental ages were obtained from the Otis norms. Since the tests were given over a period of three months it was necessary to compute mental ages to a constant date. They were computed as of September 1, 1923 the time of entrance into the sixth grade.

In most studies of test ratings of high school pupils the tests have been given in high school. The advantages of using test ratings obtained when pupils were in the sixth grade rather than in high school call for special comment.

Test norms are far more satisfactory for sixth grade children than for those of high school age. It is recognized that adequate norms are not available for children of the ages commonly found in a high school group. On the group test used in this study percentile ranks for

---

[1] Cattell, Psyche: Why Otis IQ Cannot Be Equivalent to the Stanford-Binet IQ. *Journal of Educational Psychology*, Vol. XXII, 1931, p. 599.

[2] Cattell, Psyche: IQ's and the Otis Measure of Brightness. *Journal of Educational Research*, June, 1930.

[3] Otis, Arthur S.: "Manual of Directions for Advanced Examination." World Book Company, 1929, p. 10.

children above the age of fifteen or sixteen have been worked out only by a theoretical process of extrapolation. There are very few of the children in the entire sixth grade group whose ages are subject to this disadvantage, and there are none above the age of fifteen years, one month among the sixth grade children whom we have found appearing later in the group of high school graduates.

Tests, especially group tests, which are designed to measure intelligence have been justly criticized as measuring educational achievement as well as native ability. The force of this criticism is reduced when all pupils have a comparable background of school opportunity. Pupils may fairly be compared with one another on a test given to all of them at sixth grade level, but the difference between the same individuals may be distorted appreciably by the influence of educational factors if tests are given later when some have continued on into high school while others have dropped out before reaching high school.

From the point of view of throwing light on the value of test results in educational guidance, it is desirable to consider the results of tests given at a time when they might naturally be presented if they were to be practically serviceable for advice regarding the wisdom of attempting high school work. Tests given at sixth, seventh or eighth grade level meet this requirement better than tests given later.

DATA REGARDING CHRONOLOGICAL AGES

There is available also, for this entire group, information regarding their chronological ages when tested. The information regarding ages was based on both the school record and the child's entry on the test blank. In case of discrepancy between these, further effort was made to establish the correct date of birth. Where there has not been occasion to obtain official verification the age records are subject to some degree of unreliability in case of either misrepresentation or uncertainty regarding dates of birth. All ages were computed as of September 1, 1923, the beginning of the school year during which the tests described above were given.

RESULTS—PERCENTILE RANKS

The percentile ranks of the nine hundred thirteen graduates are shown in Table III. It is apparent that the distribution in this group is considerably skewed toward the lower end. There are many more individuals with high percentile ranks than with low ones. Since an

unselected group would approximate a rectangular distribution, the selected nature of the high school graduate group is obvious.

The median percentile rank for the total group of high school graduates is 73.8, that is, 27.9 points higher than the median (45.9) for the total sixth grade group of 4184 pupils from which the graduates come. See Table I. (The fact that the median for the total group is below fifty may be explained on the basis of the number of over-age dull pupils who were repeating the grade or had repeated grades previous to the sixth.)

TABLE III.—PERCENTILE RANKS ON OTIS ADVANCED EXAMINATION GIVEN IN THE SIXTH GRADE FOR NINE HUNDRED THIRTEEN PUPILS WHO WERE GRADUATED FROM HIGH SCHOOL FIVE, SIX OR SEVEN YEARS LATER

Percentile ranks	Number graduated from high school in			Total
	1929	1930	1931	
95–100	8	64	3	75
90– 95	3	76	5	84
85– 90	8	86	6	100
80– 85	5	79	6	90
75– 80	2	78	9	89
70– 75	2	60	15	77
65– 70	1	54	14	69
60– 65	.....	52	12	64
55– 60	2	48	15	65
50– 55	.....	36	14	50
45– 50	.....	24	15	39
40– 45	1	25	11	37
35– 40	.....	19	10	29
30– 35	.....	11	4	15
25– 30	.....	10	3	13
20– 25	.....	5	1	6
15– 20	.....	2	4	6
10– 15	.....	1	.....	1
5– 10	.....	.....	2	2
0– 5	.....	2	.....	2
Total	32	732	149	913
Median	86.88	76.09	58.50	73.80
Quartile deviation	7.5	13.75	13.25	14.70

In Table III the percentile ranks of the graduates for 1929, 1930 and 1931 are shown in separate columns. The 1929 group of pupils,

who must have skipped one grade between the sixth and the twelfth grades, had no percentile ranks under forty. Their median, 86.88, is considerably higher than the medians for the other two groups. The 1930 group, which has advanced a grade a year from the time of our tests in the sixth grade, had a median of 76.09, that is ten points lower than the median of the 1929 group and seventeen points higher than that of the 1931 group. The 1931 pupils had a median percentile rank of 58.50. These are children who have in some way spent an extra year between the sixth grade, where they were tested in 1924, and their graduation from high school. Repetition of grade, loss of time from school, or change in type of school may be responsible for the extra year. It is evident that the three groups show the effect of some degree of preferential selection with respect to intelligence as measured by the Otis Examination.

In Table IV the percentile ranks of the group of high school graduates are given in larger intervals. The data show that only 1.9 per cent of the graduates had percentile ranks below twenty-five. Only 14.6 per cent had percentile ranks between twenty-five and fifty. There were 83.5 per cent that had percentile ranks above fifty, and 47.9 per cent had percentile ranks above seventy-five. That is to say, 47.9 per cent of the graduate group made test records comparable with those of the highest twenty-five per cent of an unselected group, while 83.5 per cent did as well as the upper fifty per cent of an unselected group, and only 1.9 per cent dropped as low as the lowest twenty-five per cent of an unselected group.

TABLE IV.—DISTRIBUTION OF GRADUATES IN TERMS OF NUMBER AND PER CENT OF TOTAL WITH REFERENCE TO PERCENTILE RANK ON OTIS ADVANCED EXAMINATION GIVEN IN THE SIXTH GRADE

Percentile rank	Pupils graduated from high school in						Total	
	1929		1930		1931			
	Number	Per cent	Number	Per cent	Number	Per cent	Number	Per cent
75–100	26	81.3	383	52.3	29	19.5	438	47.9
50– 75	5	15.6	250	34.1	70	46.9	325	35.6
25– 50	1	3.1	89	12.2	43	28.9	133	14.6
Below 25.....	0	0	10	1.4	7	4.7	17	1.9
Total.....	32	100	732	100	149	100	913	100

The contrast between the 1929, the 1930 and 1931 groups is again very apparent. For the lowest interval, percentile ranks from zero to twenty-five, the three groups show percentages of zero, 1.4 and 4.7 respectively, while for the highest interval, seventy-five to one hundred, their percentages are 81.3, 52.3 and 19.5 respectively.

Inspection of the distribution of percentile ranks for the entire group of graduates shows that no dead-line can be drawn below which high school graduation is never achieved. Graduates are found throughout the range of distribution of percentile ranks although the numbers are greatly diminished in the lower ranges. It is evident that the probability of successful high school completion is very slight in the case of a child with a percentile rank below twenty-five.

From such data as were available an effort was made to learn something of the personal histories and school records of the high school graduates with percentile ranks below twenty-five. Table V shows the records of all the pupils with percentile ranks below twenty-five.

Certain factors were found that should lead us to interpret with caution low ratings on group tests. We note especially foreign language handicap and reading disability. These are both factors which affect the reliability of the test results. They should be facts familiar to the teacher and the counselor in case of any particular child and accordingly not overlooked in the interpretation of a low percentile rank.

Our information is not sufficient to tell us in what proportion of cases such factors were operative. The following may serve as examples:

The two individuals with percentile ranks below five, Cases *A* and *B*, were foreign born children. One had come from Russia at the age of eleven and the other at the age of thirteen. At the time of the test they had been in this country four years and two years respectively. It is probable that the language factor was responsible for these low percentile ranks and that the test underrated the pupils' real ability.

A girl with a percentile rank of sixteen, Case *D*, attained an intelligence quotient of one hundred on an individual test. She was in the third grade at the age of eight but could not do first grade reading. Reading disability rather than limited mentality undoubtedly caused her low group test rating.

The subject of reading disability cannot be discussed here further than to emphasize its importance in connection with any use of group tests. Most group tests above the first grade level are dependent on reading ability to so great an extent that a child of average or above average mentality may rate very low on a group test if he is a poor

TABLE V.—PUPILS WITH PERCENTILE RANKS UNDER TWENTY-FIVE

Case	Percentile rank	Chronological age September 1, 1923	Mental age September 1, 1923	Curriculum
		Graduated in 1930		
A	2	14–6	8–7	General
B	2	14–4	8–5	Commercial
C	12	15–1	10–8	Commercial
D	16	12–2	9–7	General
E	19	11–9	9–7	Commercial
F	22	12–2	10–0	Industrial Arts
G	23	11–5	9–7	Commercial
H	23	11–9	9–10	Commercial
I	24	13–0	10–10	Commercial
J	24	12–0	10–1	General
		Graduated in 1931		
K	5	11–7	8–1	General
L	8	12–8	8–11	Art
M	17	12–0	9–6	General
N	17	12–0	9–7	Commercial
O	19	12–4	9–11	Industrial Arts
P	19	12–11	10–3	Commercial
Q	22	12–7	10–4	Commercial

reader. The number of poor readers and non-readers in the schools is probably greater than is realized. Wherever vocational counselors or teachers use group tests for guidance, care should be taken to have school marks in reading and to follow up low ratings on group tests with reading tests or with individual intelligence tests in cases where there is reason to question the reliability of the test rating.

Further consideration of individual cases with low percentile ranks has suggested the importance of certain other factors which may compensate to some degree for mental inadequacy. The following cases are illustrative of such factors:

The pupil with a percentile rank of twelve, who was graduated from the Commercial curriculum, was eighteen years old at the time of entering high school. She had repeated the first, second and third grades of the elementary school. One of the social agencies had been interested in her and had placed her in an excellent environment where every opportunity and incentive were given for study. Her record suggests that graduation was achieved as a result of dogged persistence and favorable external surroundings.

A girl with a percentile rank of five, Case *K*, completed high school in Summer School 1931. An individual Stanford-Binet examination given her after this

showed an Intelligence Quotient of only eighty-two but the examiner noted that her personality would be a great asset to her in whatever she might undertake. She was courteous, docile, and very anxious to please.

The cases just cited, and others for whom we have less full information, suggest that success in high school is sometimes achieved by children of rather low mental rating if there is sufficient perseverence on their part and encouragement from others for them to persist in school when they are older than their classmates and when perhaps even then they must repeat additional grades or attend summer school before finishing high school. It is worthy of note that the ages of all seventeen children with percentile ranks below twenty-five were above the median age for the entire group of graduates (eleven years, 2.8 months) and in several cases were very much above this. There are suggestions also of the presence of favorable personality traits in some cases which have doubtless facilitated their school adjustments.

There remain certain cases whose records reveal no exceptional factors accounting for their success in the face of low mental ratings. We may note the following as representative of these:

A girl with a percentile rank of nineteen, Case *E*, completed the commercial curriculum in 1930. She spoke English with ease though Italian was the language of the home. She was fifteen when she entered high school. An individual Stanford-Binet test gave her an intelligence quotient of eighty-five. This pupil was apparently adequately rated by the group test and as far as our information shows did not have her record lowered by any such factors as language disability or poor reading ability.

A boy with a percentile rank of twenty-two, Case *F*, was graduated in 1930 from the Industrial Arts curriculum. He had attended Cincinnati schools throughout his school career and had repeated only one grade before entering the sixth grade. He attended a regular high school for three years and then went to Night High School and was graduated there.

We may summarize the situation regarding low percentile ranks for the pupils graduating in 1929, 1930 and 1931. In the 1929 group no pupils had percentile ranks under forty. In the 1930 group the lowest percentile rank that was apparently an adequate rating for the individual was nineteen, Case *E*. The 1931 group had two pupils with percentile ranks below ten (Cases *K* and *L*).

RESULTS—PERCENTILE RANKS WITH REFERENCE TO CURRICULA

The Cincinnati program of studies offers seven different curricula each of which prepares for admission to college provided a certain selection of variables or electives is taken. Two of the curricula,

namely General and Classical, are distinctly college preparatory. The remaining five, Commercial, Industrial Arts, Household Arts, Music and Art, make college entrance possible but do not emphasize it.

Table VI gives the distribution of percentile ranks for the pupils in each curriculum. The curriculum with the highest median percentile rank, 93.75, is the Classical but since this includes only seven cases the median is not reliable. Of the curricula that include enough cases to be representative the General curriculum has the highest median, that is 77.33. The median for the Commercial curriculum is 61.67.

TABLE VI.—PERCENTILE RANKS ON OTIS ADVANCED EXAMINATION GIVEN IN THE SIXTH GRADE

1929, 1930 and 1931 High School Graduates Separated for Curricula

Percentile rank	General	Classical	Commercial	Industrial arts	Household arts	Music	Art	Grand total
95–100	67	3	4	1	.....	.....	.....	75
90– 95	70	2	7	.....	2	3	.....	84
85– 90	78	1	12	4	3	2	.....	100
80– 85	68	1	13	5	1	2	.....	90
75– 80	72	.....	2	11	3	.....	1	89
70– 75	55	.....	14	5	.....	2	1	78
65– 70	37	.....	15	8	6	3	.....	69
60– 65	51	.....	9	.....	1	2	1	64
55– 60	39	.....	19	3	1	3	.....	65
50– 55	29	.....	10	6	3	2	.....	50
45– 50	20	.....	10	3	4	.....	2	39
40– 45	22	.....	7	3	2	.....	3	37
35– 40	15	.....	9	2	2	.....	1	29
30– 35	6	.....	4	1	3	.....	1	15
25– 30	9	.....	2	2	.....	.....	.....	13
20– 25	1	.....	4	1	.....	.....	.....	6
15– 20	2	.....	3	1	.....	.....	.....	6
10– 15	.....	.....	1	.....	.....	.....	.....	1
5– 10	1	.....	.....	.....	.....	.....	1	2
0– 5	1	.....	1	.....	.....	.....	.....	2
Total	643	7	146	56	31	19	11	913
Median	77.33	93.75	61.67	68.75	62.50	69.17	44.17	73.80
Quartile deviation	13.47	4.17	15.50	13.70	15.58	13.02	11.25	14.67

For the Industrial Arts, Household Arts, and Music curricula also the medians run slightly lower than for the General curriculum

although it should be noted that they have no representatives falling so low in percentile rank as those with the lowest ratings in the General and the Commercial groups. The number of the present group in the Art curriculum is too small to warrant more than casual notice.

RESULTS—MENTAL AND CHRONOLOGICAL AGES

Table VII gives a distribution of mental ages for the nine hundred thirteen high school graduates based upon the Otis Advanced Examina-

TABLE VII.—MENTAL AGES ON OTIS ADVANCED EXAMINATION GIVEN IN THE SIXTH GRADE 1923–1924 COMPUTED TO SEPTEMBER 1, 1923 FOR NINE HUNDRED THIRTEEN PUPILS WHO WERE LATER GRADUATED FROM HIGH SCHOOL

Mental age at time of entrance to sixth grade	Number of pupils graduated from high school in			Total
	1929	1930	1931	
Above 18 years	1	29	.....	30
17 yr. 6 mo.–18 yr	.....	2	1	3
17 yr. –18 yr. 6 mo	.....	4	.....	4
16 yr. 6 mo.–17 yr	3	4	1	8
16 yr. –16 yr. 6 mo	6	22	2	30
15 yr. 6 mo.–16 yr	1	35	.....	36
15 yr. –15 yr. 6 mo	.....	50	4	54
14 yr. 6 mo.–15 yr	7	50	4	61
14 yr. –14 yr. 6 mo	6	73	8	87
13 yr. 6 mo.–14 yr	4	75	9	88
13 yr. –13 yr. 6 mo	1	88	17	106
12 yr. 6 mo.–13 yr	2	66	15	83
12 yr. –12 yr. 6 mo	.....	60	13	73
11 yr. 6 mo.–12 yr	1	62	20	83
11 yr. –11 yr. 6 mo	.....	44	17	61
10 yr. 6 mo.–11 yr	.....	35	22	57
10 yr. 10 yr. 6 mo	.....	17	8	25
9 yr. 6 mo.–10 yr	.....	12	5	17
9 yr. – 9 yr. 6 mo	.....	2	1	3
8 yr. 6 mo.– 9 yr	.....	1	1	2
8 yr. – 8 yr. 6 mo	.....	1	1	2
Total	32	732	149	913
Median	14 yr. 7.7 mo.	13 yr. 4.5 mo.	11 yr. 11.9 mo.	13 yr. 2.8 mo.
Quartile	1 yr. 1 mo.	1 yr. 3.4 mo.	1 yr. 1.7 mo.	1 yr. 3.7 mo.

tion given in the sixth grade. The mental ages were computed to a constant date, September 1, 1923. Since the mental age is a measure of maturity regardless of brightness, these data give an indication of the mental maturity of the graduates when they entered the sixth grade.

It is commonly assumed that a mental age of eleven or over is normal for sixth grade and norms for educational tests imply an even higher mental age. In Table VII there are one hundred six pupils or 11.6 per cent with mental ages under eleven. Twenty-four pupils or 2.6 per cent have mental ages below ten. Some of the lowest mental ages may be accounted for as unreliable in view of factors already discussed in connection with percentile ranks but even so there remains an appreciable number with mental ages below eleven for whom we have no such explanation. The per cent below eleven, however, is small as compared with 72.62 per cent with mental ages of twelve or over.

TABLE VIII.—CHRONOLOGICAL AGES AT TIME OF ENTRANCE INTO SIXTH GRADE FOR NINE HUNDRED THIRTEEN PUPILS WHO WERE LATER GRADUATED FROM HIGH SCHOOL

Chronological age at time of entrance to sixth grade	Number of pupils graduated from high school in			Total
	1929	1930	1931	
15 yr. –15 yr. 6 mo.	.....	1	.....	1
14 yr. 6 mo.–15 yr.	.....	1	.....	1
14 yr. –14 yr. 6 mo.	.....	1	.....	1
13 yr. 6 mo.–14 yr.	1	3	1	5
13 yr. –13 yr. 6 mo.	.....	7	2	9
12 yr. 6 mo.–13 yr.	.....	12	11	23
12 yr. –12 yr. 6 mo.	2	53	14	69
11 yr. 6 mo.–12 yr.	8	159	31	198
11 yr. –11 yr. 6 mo.	13	257	60	330
10 yr. 6 mo.–11 yr.	6	170	21	197
10 yr. –10 yr. 6 mo.	1	49	6	56
9 yr. 6 mo.–10 yr.	1	17	1	19
9 yr. – 9 yr. 6 mo.	.....	1	1	2
8 yr. 6 mo.– 9 yr.	.....	1	1	2
Total	32	732	149	913
Median	11 yr. 3.2 mo.	11 yr. 2.4 mo.	11 yr. 4.0 mo.	11 yr. 2.8 mo.
Quartile deviation	4.13 months	4.99 months	4.74 months	4.92 months

A comparison of Table VII with Table VIII (Distribution of Chronological Ages) confirms the conclusion based on the percentile rank distribution that the graduates are a selected group with reference to intelligence. The median mental age, thirteen years 2.8 months, is two years higher than the median chronological age of eleven years 2.8 months. The distribution of mental ages has a wider range than that of chronological ages. The highest chronological age at time of entrance to the sixth grade was fifteen, whereas there were thirty pupils with mental ages above eighteen. Only 1.9 per cent of the group had chronological ages of thirteen years or more while 55.5 per cent had mental ages of thirteen or more. Thirty and two-tenths per cent had chronological ages below eleven while only 11.6 per cent had mental ages below eleven.

SUMMARY

1. It is evident that the high school graduates are a selected group from the point of view of intelligence as measured by a group test administered in the sixth grade. This is shown first by the fact that the median percentile rank of the group is 73.80 as contrasted with 45.87, the median percentile rank of the entire sixth grade group tested in 1924. It is also shown by the fact that 47.9 per cent of this group made records comparable with those of the highest twenty-five per cent of an unselected group, that 83.5 per cent did as well as the upper fifty per cent of an unselected group, and that only 16.5 per cent dropped as low as the lowest fifty per cent of an unselected group.

2. Although the range of percentile ranks for the graduate group was from below five to one hundred, low ratings should be interpreted with caution. Such factors as language handicap and poor reading ability may produce low group test ratings for pupils of average intelligence. Where vocational and educational counseling is carried on on the basis of group test ratings, individual case studies should be made in order to check up on possible language handicap or reading disability.

3. Pupils with percentile ranks under twenty-five may achieve graduation but the probability of this is slight. Only 1.9 per cent of the graduates made ratings as low as this. The pupils with low percentile ranks who did achieve graduation tended to do so at considerable expense of time. Repetition of grades either before or after the time of the sixth grade test was a noticeable tendency.

4. The percentile ranks for the three groups of graduates, those graduated in 1929, in 1930 and in 1931, show some degree of preferential selection with respect to intelligence as measured by the Otis Advanced Examination. The median for the 1929 group was 86.88, that for the 1930 group was 76.09 and that for the 1931 group was 58.50.

5. The comparison of median percentile ranks for pupils graduated in each of the seven curricula show no outstanding differences except for the General and Classical curricula. The Classical curriculum had the highest median but included very few cases. The General curriculum had the highest median of any curriculum with enough cases to be significant. The medians for the other curricula were approximately at the same level.

6. The mental ages (figured to September 1, 1923, time of entrance into Grade VI) show that the high school graduates are a selected group from point of view of mental maturity. The median mental age thirteen years, 2.8 months is two years higher than the median chronological age of eleven years, 2.8 months. Only 1.9 per cent of the pupils had chronological ages above thirteen whereas 55.5 per cent had mental ages above thirteen. There is, however, a wide range of mental ages from eight years to above eighteen years. Very few overage pupils were in the group of graduates. Only twelve per cent of the group entered the sixth grade with ages above twelve. Thirty per cent were under eleven years old upon entering the sixth grade. Overage pupils are much less likely to achieve graduation than underage pupils or pupils at age for their grade.

A DIAGNOSTIC STUDY OF INDIAN HANDWRITING

ELIZABETH WEISSER

University of Denver

Our purpose in this study was to compare the various phases of handwriting as slant, alinement, quality of line, letter form, and spacing of white children, full-blood Indian children, and mixed-blood Indian children. Through this study we wished to determine whether differences, if any existed, were racial or not.

A study has been recently made by Thomas R. Garth comparing white children, full-blood Indian children, and mixed-blood Indian children in regard to legibility score and speed. I do not know of any other study along this line of investigation.

In this study the handwriting specimens of two hundred and sixty whites, six hundred and three full-blood Indians, and one hundred and eighty four mixed-blood Indians were used. The specimens were obtained at the Albuquerque, New Mexico, Indian School, the Chilocco, Oklahoma, Indian School, and in the schools in and near Denver, Colorado. Only the specimens from the fourth through the eighth grades were used in this study. To obtain these specimens the children were asked to write a verse of "Mary had a little lamb" as many times as they could in five minutes. They were urged to write in their usual manner and with their customary speed. These specimens were scored first by Thorndike's Handwriting Scale to obtain the legibility score and then by Freeman's Diagnostic Scale to determine the score for slant, alinement, quality of line, letter form, and spacing. The number of letters written per minute was also determined. The scoring of these papers was done by more than one person, the average being taken for the final result in order to do away with the personal equation as much as possible in this subjective type of scoring.

By studying Table I we find that either class of Indians sometimes exceeds and sometimes falls below the white medians but there seems to be comparatively little difference as the medians vary only slightly. In the case of the correlations there is again comparatively little difference between whites and Indians. In the case of the full-blood Indians and the mixed-blood Indians we can easily see that the two classes vary little, neither maintaining its relative position throughout the grades. The whites seem usually to exceed the Indians in the upper grades and fall below them in the lower grades. This condition is

TABLE I

Grade	Total number	Age		Slant		Alinement		Quality of line		Letter form		Spacing		Total diagnostic score			
		Median	Q	Median	Q	Median	Q	Median	Q	Median	Q	Median	Q	My median	Q	Freeman's norms	Per cent overlap
Whites																	
4	38	10.14	.66	2.81	.66	2.3	...	2.57	...	2.79	.70	2.93	.66	14.5	3.6	14.5	50
5	58	10.84	.64	2.82	.75	2.42	...	2.29	...	2.33	...	2.53	...	13.78	4.0	16	29
6	61	12.09	.61	3.47	.51	2.89	.81	3.02	.69	3.43	.75	3.46	.75	18.6	3.42	18	57
7	58	12.57	.53	3.50	.71	3.14	.83	3.22	.82	3.05	.66	3.35	.74	18	3.83	20	31
8	45	13.62	.83	3.71	.70	3.66	.80	3.59	.75	3.38	.64	3.63	.74	20.37	4.24	21	49
T	260	12.04	1.04	3.31	.77	2.87	.87	2.90	.75	3.01	.75	3.23	.72	16.78	3.81		

TABLE I.—*Continued*

Grade	Total number	Age		Slant			Alinement			Quality of line			Letter form			Spacing			Total diagnostic score			
		Median	Q	Median	Q	Per cent overlap	Median	Q	Per cent overlap	Median	Q	Per cent overlap	Median	Q	Per cent overlap	Median	Q	Per cent overlap	Median	Q	Per cent overlap my median	Per cent overlap Freeman's norm
Full-blood Indians																						
4	37	14.58	2.17	2.5	...	41	2.32	...	51	2.29	...	41	2.34	...	27	2.19	...	39	12.5	3.16	35	35
5	64	15.75	1.67	2.58	.60	41	2.62	.66	59	2.81	.65	71	2.56	...	61	2.74	.52	58	14.00	3.34	55	33
6	129	14.98	1.64	2.45	...	23	2.5	...	34	2.79	.80	42	2.38	...	15	2.64	.69	22	13.5	3.67	15	18
7	191	15.45	1.31	3.1	.69	34	2.84	.61	39	3.09	.67	45	2.94	.72	44	3.14	.70	21	16.7	3.47	41	29
8	182	16.64	1.20	3.41	.77	39	3.42	.70	39	3.54	.74	47	3.26	.76	43	3.76	.87	55	19.1	3.82	35	31
T	603	15.97	1.44	2.97	.77	39	2.86	.73	50	3.09	.72	57	2.80	.75	40	3.09	.81	46	16.8	3.73	46	
Mixed-blood Indians																						
4	14	16.2	1.2	3.14	.74	64	3.0	...	71	2.50	...	50	2.43	...	35	2.20	...	28	13.32	3.59	43	43
5	35	14.69	1.25	3.03	.60	60	2.69	.63	65	3.14	.57	80	2.44	...	57	2.95	.75	13	15.0	2.7	65	37
6	35	16.75	1.34	3.12	.76	37	2.75	...	45	2.77	...	43	2.75	...	26	2.71	...	26	15.0	4.17	29	31
7	60	16.07	1.17	3.32	.90	44	2.63	.57	28	2.60	...	32	2.86	.47	43	2.83	.88	35	16.0	2.99	33	18
8	40	16.73	.84	3.25	.84	48	3.06	.59	25	3.33	...	38	3.78	.60	72	3.47	...	45	16.32	3.34	23	18
T	184	16.22	1.33	3.18	.77	45	2.78	.63	46	2.95	.73	52	2.70	.68	34	2.93	.80	41	15.56	3.20	40	

probably due to the fact that in the lower grades the Indians are much older than the whites, while in the upper grades the difference in age is not so marked although the Indians are somewhat older than whites

TABLE II.—CORRELATIONS

	Slant	Alinement	Quality of line	Letter form	Spacing
Whites					
Legibility score	.72(.02)	.75(.02)	.75(.02)	.79(.03)	.51(.03)
Speed	.10(.04)	.21(.04)	.15(.04)	.04(.04)	.08(.04)
Age	.28(.04)	.33(.04)	.27(.04)	.17(.04)	.27(.04)
Grade	.39(.04)	.39(.04)	.34(.04)	.24(.04)	.31(.04)
Slant	.....	.75(.02)	.75(.02)	.71(.02)	.57(.03)
Alinement	.....	.....	.83(.01)	.73(.01)	.58(.03)
Quality of line	.....	.....	.....	.79(.02)	.60(.03)
Letter form	.....	.....	.....	.....	.60(.03)
Full-blood Indians					
Legibility score	.57(.02)	.67(.02)	.71(.01)	.69(.01)	.63(.01)
Speed	.10(.03)	.14(.03)	.38(.02)	.06(.02)	.21(.03)
Age	.19(.03)	.07(.03)	.04(.03)	.07(.03)	.02(.03)
Grade	.29(.03)	.36(.03)	.34(.02)	.29(.03)	.38(.03)
Slant	.....	.68(.01)	.65(.01)	.63(.02)	.51(.02)
Alinement	.....	.....	.80(.01)	.73(.01)	.64(.02)
Quality of line	.....	.....	.....	.42(.02)	.71(.01)
Letter form	.....	.....	.....	.....	.62(.02)
Mixed-blood Indians					
Legibility score	.54(.04)	.59(.03)	.69(.03)	.58(.04)	.35(.04)
Speed	.09(.05)	.02(.05)	.12(.05)	.12(.05)	.14(.05)
Age	.05(.05)	−.07(.05)	−.01(.05)	.03(.05)	−.07(.05)
Grade	.12(.05)	.08(.05)	.13(.05)	.21(.04)	.37(.04)
Slant	.....	.62(.03)	.60(.03)	.54(.04)	.49(.04)
Alinement	.....	.....	.72(.02)	.64(.03)	.61(.03)
Quality of line	.....	.....	.....	.60(.03)	.73(.02)
Letter form	.....	.....	.....	.....	.60(.03)

The number in parenthesis is the probable error of the correlation and is always a plus or minus number.

in every grade. This difference in age might account for the Indians exceeding the whites in the lower grades since younger children do not have the same muscular control as older children.

It is impossible to say that any differences that exist between whites and Indians in handwriting are due to race since the age element was not controlled. It is possible that teaching or environmental conditions might be the cause of the differences. This study does not indicate any differences that might be attributed to race since in some cases Indians seem to exceed white children in their ability to write. It is not probable that any differences that exist are due to an inferiority in the Indian race.

After the specimens had been scored correlations were found for legibility score with each of the phases of handwriting, also speed, age, and grade with each phase and finally each phase with every other phase. For each grade the average, standard deviation, median and quartile deviation for each phase of handwriting were found and also the percent overlap for the full-bloods and the mixed-blood Indians over whites. The total diagnostic score, that is the sum of the scores for the five phases of handwriting, was also found and the median and quartile deviation found for each grade. Professor Freeman of the University of Chicago furnished me with norms for each grade for the total diagnostic score. Using these norms with my own for whites I found the per cent overlap for the two classes of Indians over whites.

Table I is in three sections: First, whites; second, full-blood Indians; third, mixed-blood Indians. It shows medians quartile deviations and per cent overlap of Indians over whites for age, slant, alinement, quality of line, letter form, spacing, and total diagnostic score for grades four through eight and also for the total number.

Table II shows the correlation results obtained for whites, full-blood Indians, and mixed-blood Indians. The correlations may be found by referring to the table.

BOOK REVIEWS

E. L. THORNDIKE. *The Fundamentals of Learning.* Pp. XVII + 638. New York: Bureau of Publications, Teachers College, Columbia University, 1932.

RAYMOND H. WHEELER and F. THEODORE PERKINS. *Principles of Mental Development; A Textbook in Educational Psychology.* Pp. XXVI + 529. New York: Thomas Y. Crowell Co., 1932.

These two books, fundamentally different in their outlook and conclusions, may, nevertheless, be usefully reviewed together. Thorndike is the father of the reaction-hypothesis; Wheeler and Perkins are representatives of the Gestalt group of psychologists.

When Thorndike published his Cornell lectures under the title of Human Learning, he promised later to supply the details of the evidence on which his conclusions were based. This promise has been fulfilled by the publication of "The Fundamentals of Learning." Let it be said at the outset that the new volume is not easy to read. It is packed full of evidence, chiefly in the form of 153 tables. Thorndike's subjects were human, namely, students and others in and around Teachers College. Through them he subjected his laws of learning to a very severe and critical testing and analysis. On the whole his previous work stands up remarkably well. Those of us who have followed his work closely will note the modifications that must be made in the light of the new evidence. The law of exercise as previously formulated needs re-drafting. It is now seen that repetition without belongingness (Thorndike's horrible word) or sequence is impotent and useless. Repetition *qua* repetition produces no beneficial effects. On the other hand, the law of effect comes out stronger than ever and assumes an unchallengeable first place in the series. Rewards prove to be more effective than punishment; in fact the latter is found to have singularly little effect on the course of learning. Learning is not, as the gestaltists claim it to be, an all-or-none affair, but a steady accumulation of right responses with gradualness an essential factor. Certain aspects of learning, which Thorndike terms belongingness, availability, impressiveness (vividness) and identifiability, are examined at length and declared to be important. Attentiveness, unless accompanied by belongingness, is an ineffective factor in learning. From the results of his experiments, one may conclude that Thorndike remains satisfied both with the laws of learning and the reaction-hypothesis and that

with the above modifications they may stand as previously given. Like all his writings this new work demands a critical reading; a task which once performed amply repays the serious student.

Turning now to Wheeler and Perkins' book, we find the most serious challenge to the teachings of Thorndike and his school that has yet been made. The authors have tried to write a textbook in Educational Psychology from the strictly gestalt point of view. They have cited a very large number of experiments, including the very important ones by Coghill and Lashley, which are interpreted as supporting their case. Every principle or law put forward by believers in the reaction-hypothesis is strongly challenged and stoutly contradicted. In very large capitals we learn that "The very essence of learning is NOT REPEATING A PERFORMANCE, BUT MAKING A NEW ONE." The laws of human nature are declared to be those of (1) field properties, (2) derived properties, (3) determined action, (4) individuation, (5) field genesis, (6) least action, (7) maximum work, and (8) configuration. These laws are shown to be active in both animal and human learning. Illustrations of the operation of the laws occur frequently throughout the text. More than half the space is devoted to the problem of learning in one aspect or another and the authors are to be congratulated on the mass of experimental evidence they have accumulated. The reviewer is grateful for having his attention directed to more than a dozen studies that he had overlooked in his previous reading.

Having pointed out the excellences, it is now my painful duty to note some defects. Some of these have undoubtedly crept in through hurried composition, but some are inherent in the structure of the text. On page 68 we find (in italics): "Thus the movements of twelve pairs (for each eye has six pairs) of muscles are simultaneously given direction in terms of the conditions for least action." This is obviously a slip of the pen, but what can be said of the following, which is used as an explanation of the creation of capacity or intelligence?

> "The real function of environment in the production of an organism may be understood with the aid of an analogy. A very sluggish stream of water flows through a canyon. There is no power in the stream; it can achieve nothing. But an engineer builds a dam across it, raising its level two hundred feet. Now the stream rushes over the top of the dam and at the bottom turns mighty engines that supply power for a radius of many miles. *A potential has been raised by elevating the stream. Now there is achievement.*" (P. 165.)
>
> "Now we know the function of Education. Its aim is to raise the energy of the growing organism to higher potentials through stimulation and motivation.

It does more than permit intelligence to express itself; *it creates intelligence*, just as building a dam creates water power." (P. 166.)

This view of education seems to be as naive as the hydrodynamics used to illustrate it. Yet the authors in the succeeding chapter, "The Measurement of Intelligence" give as conventional an account of the subject, including the statistical treatment of results, as a Pintner or a Terman would give. Nor are they always fair in the selection of examples. They quote Freeman's investigation of the intelligence of adopted children, which supports their viewpoint, and neglect Burks', which doesn't. And on what evidence do they base the statement that the average IQ of children in general is 81? (P. 169.) We always thought it was one hundred. Ebbinghaus first subjected the process of forgetting to experimental (not laboratory) control in 1879–1880, not in 1885 as stated on p. 399. The authors would be well-advised to give their work a careful rereading and to remove the rather large number of inaccuracies they have permitted to creep in.

Notwithstanding the above criticisms, this work of Wheeler and Perkins is immensely stimulating. All the disciples of Thorndike should give it an especially careful reading. They will certainly get a number of shocks.

PETER SANDIFORD.
University of Toronto.

JAMES L. MURSELL and MABELLE GLENN. *The Psychology of School Music Teaching*. Pp. V + 378. New York: Silver, Burdett and Co., 1931.

This is a general treatment of school music education: the musical nature of the child, the need for the appreciation aim, problems of executant music and measures of musical capacity. The approach to technique is through the voice. Formal drill has no place in the educative plan. The viewpoint naturally is modern, stressing the will-to-learn rather than the doctrine of specific abilities. Technique becomes artistic playing or singing instead of mere mechanics. The correlation between psychological principles and classroom problems should help teachers of school music in systematizing procedures, and basing them upon relatively safe fundamentals. The questions at the ends of chapters lend the work text-book value, and the bibliography is ample for much valuable collatoral reading.

So long as the viewpoint of the authors, namely, school music teaching, is remembered, their conclusions are sound, for the object here intended is music appreciation, a flexible and very adaptable aim.

So soon as we demand vocal or instrumental performances of musically artistic merit the aim changes considerably and with it the teaching procedure. The authors fail to stress this difference, particularly as to the drill necessary for any performance, and, as a result, treatment of topics leading over into the field of more specialized music is weaker than that concerning general classroom situations.

The insistence of the authors upon the teachers of school music being truly musical personalities is as significant as it is excellent. However, the musical training of such teachers will always remain incomplete and relatively unsatisfactory until teachers and colleges realize that the school-music teacher's ability to sing a tune with good quality and in correct intonation is more important that a course in science or mathematics, and that a feeling for real musical values is to be preferred to a training in history of education.

OTTO ORTMANN.

Peabody Conservatory of Music, Baltimore.

A. CAMPBELL GARNETT. *The Mind in Action, A Study of Motives and Values.* Pp. 226. New York: D. Appleton & Co., 1932.

The very recent publication of this volume indicates that the doctrine of instincts is not dead, despite the numerous funeral rites that have been pronounced over it by leading psychologists during the last ten years. Dr. Garnett presents a treatise on motivation of which Instinct is the corner stone. Says the author "We find then that, in addition to the appetites, or recurrent wants due to internal stimuli, man has the following instincts, or occasional wants due to the stimulating effect of an external, significant, situation: nutrition, sex, sleep, escape, pugnacity, repulsion, curiosity, acquisition, construction, hunting, parental care, appeal, gregariousness, self-display, leadership and submission." Such a list is barely distinguishable from that of William James in 1890.

Like most upholders of an instinct doctrine, Dr. Garnett does not take the trouble to show that his "instincts" are genuinely native. There is no mention in his book of the modern work in child study. James is quoted seven times in the book, McDougall nine times. Watson and Gessell are not even mentioned. All in all, the volume will delight those who studied William James when in school, and who have not changed their minds since.

LAURANCE F. SHAFFER.

Carnegie Institute of Technology.

EDWIN A. KIRKPATRICK. *The Sciences of Man in the Making.* Pp. XV + 396. New York: Harcourt, Brace & Co., 1932.

Dr. Kirkpatrick, in this volume, has made an excellent job of a most difficult task. Not only does he cover a vast field, but he also presents this vast material in a most readable style. The book is a compendium of valuable information culled from the sciences of man, significant both as a synthesis for the specialist and as general information for the layman. Books of this sort are needed, and one hopes that Dr. Kirkpatrick's volume will find many readers. A most welcome feature of the book consists of selections from typical researches illustrating scientific method in the various fields covered by the book.

MAX SCHOEN.

The Carnegie Institute of Technology.

ALICE C. BRILL AND MAY PARDEE YOUTZ. *Your Child and His Parents.* Pp. XII + 339. New York: D. Appleton and Co., 1932.

A series of lessons in child life purposed to serve as a basic text for child-study groups is "Your Child and His Parents." Ten of the eleven chapters are devoted to a description and discussion of child problems. Discipline and Imagination rate two chapters each. Play Life, Fear, Habit Formation, Heredity and Environment, Sex Education and the Child's Use of Money are the topics that rate one chapter. The last chapter deals with methods of preparing materials and conducting study groups. Appendices contain list of relevant institutions, magazines and publishers.

This book is not, as the title may suggest, a treatment of the dynamics of parent-child relationships. In fact, references to and quotations from, the writings of mental hygienists and social psychologists who are known for their contributions to our understanding of personality adjustments of children are conspicuous by their absence. Dialogues, examples and other fragmentary illustrations are included but comprehensive case studies are not. The book contains many suggestions for assignments which should be helpful to uninformed leaders.

H. MELTZER.

Psychiatric Child Guidance Clinic, St. Louis, Missouri.

C. H. STOELTING CO.

MANUFACTURERS—PUBLISHERS—IMPORTERS—EXPORTERS

Apparatus, Tests

AND

Supplies

FOR

PSYCHOLOGY, PSYCHOMETRY, PSYCHOTECHNOLOGY,
PSYCHIATRY, NEUROLOGY, ANTHROPOLOGY,
PHONETICS, PHYSIOLOGY,
AND
PHARMACOLOGY

424 N. Homan Ave., Chicago, Ill., U.S.A.

THE ACTIVITY MOVEMENT

By CLYDE HISSONG

In an attempt to overcome the weaknesses of the traditional school organization many progressive schools have developed new programs. These programs are so similar in character that collectively the changes have been referred to as the activity movement. This movement has claimed the center of the educational stage for a length of time sufficient to have engendered widespread interest in its outcomes and in its basic philosophy.

In Doctor Hissong's study an attempt has been made to discover the principles underlying the present activity movement, to determine the influence of traditional concepts in shaping the trends of the movement, and to see if in the light of the present knowledge of the child and his relation to his environment the movement rests upon a justifiable basis.

$2.00 plus 10¢ postage.

WARWICK AND YORK

Publishers BALTIMORE

University
Research Monographs

No. 1. The Value of Homogeneous Grouping, *by* T. Luther Purdom..... $2.00 + 8¢

No. 2. Contemporary Municipal Government in Germany, *by* Bertram W. Maxwell... $2.50 + 8¢

No. 3. The American Normal School, *by* Vernon L. Mangun..... $3.50 + 12¢

No. 4. State Control of Secondary Education, *by* O. L. Troxel..... $2.50 + 10¢

No. 5. Variations in Development and Motor Control in Goiterous and Non-goiterous Adolescent Girls, *by* Louise A. Nelson..... $2.75 + 10¢

No. 6. Vocational Education of Junior College Grade, *by* G. Vernon Bennett..... $2.75 + 12¢

No. 7. The Pedagogical Value of the True-false Examination, *by* A. W. Cocks..... $2.50 + 8¢

No. 8. Art Education in Elementary Schools, *by* Fredrik V. Nyquist..... $2.00 + 8¢

No. 9. Suffrage and Its Problems, *by* Albert J. McCulloch..... $2.50 + 8¢

No. 10. The Rise of the Public High School in Connecticut, *by* Silas Hertzler..... $2.75 + 12¢

No. 11. A Determination of Generalizations Basic to the Social Studies Curriculum, *by* Neal Billings. $3.00 + 12¢

No. 12. The School in American Literature, *by* Richard Allen Foster..... $2.50 + 10¢

No. 13. Anglo-German Imperialism in South Africa: 1880–1900, *by* Raymond Walter Bixler. $2.75 + 10¢

WARWICK AND YORK, Inc.,

Publishers, Baltimore

THE MAPLE PRESS COMPANY, YORK, PA.

K, PA.

METRIC
INCHES
C2399